Independent summative evaluation of the Fleming Fund, 2017–26

Final report

Authors: Tim Shorten, Lamiaa Shehata, Giada Tu Thanh, Ruth Sherratt, Ankur Gupta-Wright, Syed Abbas, Rebecca Maudling, Marco Maragno, Archie Drake, Jennifer Armitage, Cheryl Brown, Paul Balogun

Date: June 2026

Acknowledgements

This Phase 2 summative evaluation was produced by the core evaluation team comprising Tim Shorten, Lamiaa Shehata, Giada Tu Thanh, Ruth Sherratt, Ankur Gupta-Wright, Syed Abbas, Rebecca Maudling, Marco Maragno, Jon Cooper, Archie Drake, Jennifer Armitage, Cheryl Brown and Paul Balogun. Country-level data collection for this summative report was undertaken by Aminah Rajput (Pakistan), Ozioma Nwagwu (Nigeria), Pippa Page and Eric Ogola (Kenya), Anila Shafa and Mega Raj Banjara (Nepal), and Ignacio Torrano (Indonesia). Leadership and support from Itad was provided by Sarah Lamb, Anita Pavic, Christine Martin, Hanifa Ali, Bikash Gyawali and Tahmina Begum. See Annex 3.3 (Vol. II) for detailed roles.

Thanks are also due to the Department of Health and Social Care Fleming Fund Project Team, the Fleming Fund Management Agent Team at Mott MacDonald for extensive interactions during the course of data collection and consideration of the implications of the analysis, and to all who took part in our key informant interviews.

Disclaimer

The views expressed in this report are those of the evaluators. They do not represent those of the UK Department of Health and Social Care or of any of the individuals and organisations referred to in the report. 'Itad' and the tri-colour triangles icon are a registered trademark of ITAD Limited.

Suggested citation

Itad (2026) 'Independent summative evaluation of the Fleming Fund, 2017-26'. Brighton: Itad.

Copyright

Contents

List of figures

List of boxes

List of tables

List of acronyms

AH – Animal health

AMR – Antimicrobial resistance

AMROH – Antimicrobial Resistance and One Health, including Animal Health, the Environment and Practitioner Engagement (Regional grant)

AMU – Antimicrobial use

AMC – Antimicrobial consumption

AMRCC - Antimicrobial Use Coordination Committee

AMS - antimicrobial stewardship

ANIMUSE – Animal Antimicrobial Use Surveillance (WOAH global database)

AST – Antimicrobial susceptibility testing

CE – Clinical engagement

CG – Country grant

CIS – Country Investment Strategy

COVID-19 – Coronavirus disease 2019

DAC – Development Assistance Committee

DG – Direct grant

DHSC – Department of Health and Social Care (UK)

EQA – External Quality Assessment

EQ – Evaluation Question

FCDO – Foreign, Commonwealth and Development Office

FDC – Fixed-dose combination

GBD – Global Burden of Disease

G&E – Gender and equity

GEAR-UP – Gender, Equity and AMR Use Programme

GLASS – Global Antimicrobial Resistance and Use Surveillance System (WHO)

GRAM – Global Research on Antimicrobial Resistance

HH – Human health

HLM – High-Level Meeting

HR – Human resources

IHME – Institute for Health Metrics and Evaluation

ICAI - Independent Commission for Aid Impact

IPC – Infection Prevention and Control

IO – Intermediate Outcome

KII – Key informant interview

KPI – Key performance indicator

LMIC – Low- and middle-income country

MA – Management Agent

MAAP – Mapping Antimicrobial Resistance and Antimicrobial Use Partnership

NAP – National Action Plan (on Antimicrobial Resistance)

NRL – National Reference Laboratory

ODA – Official Development Assistance

OECD – Organisation for Economic Co-operation and Development

OHHLEP – One Health High Level Expert Panel

OH – One Health

P&MT - Planning & Monitoring Tool

PAHO – Pan American Health Organization

PEA – Political Economy Analysis

PPS – Point Prevalence Survey

RG – Regional grant

RF – Results Framework

SEP – Sustainability and Exit Plan

SOP – Standard operating procedure

TADEU – Technical Assistance for Development of Economic Understanding (Regional Grant)

TACE – Technical Assistance for Clinical Engagement (Regional Grant)

TAG – Technical Advisory Group

TGHN - The Global Health Network

ToC – Theory of Change

TrACSS - Tracking AMR Country Self-Assessment Survey

TWG – Technical Working Group

UMIC - Upper-middle income countries

VfM – Value for money

WGS – whole-genome sequencing

WHO – World Health Organization

WOAH – World Organisation for Animal Health (formerly OIE, Office International des Epizooties)

Glossary

Active surveillance

The systematic collection of AMR or AMU data through planned sampling of defined populations (human or animal), rather than relying on samples generated through routine care. In the Fleming Fund, active surveillance was used primarily in the animal health sector, particularly in food producing animals.

Antimicrobial consumption (AMC)

Aggregate data on the total volume of antimicrobials consumed within a country or sector, usually derived from import, sales, procurement or reimbursement records. AMC is often used as a proxy for antimicrobial use where detailed AMU data are unavailable.

Antimicrobial resistance (AMR)

The ability of microorganisms (such as bacteria) to survive exposure to antimicrobial medicines that would normally kill them or inhibit their growth, making infections harder to treat and increasing the risk of spread, severe illness and death.

Antimicrobial resistance surveillance

The continuous and systematic collection, analysis and interpretation of data on antimicrobial resistance patterns, used to inform public health action, clinical practice, policy and global monitoring.

Antimicrobial susceptibility testing (AST)

Laboratory testing used to determine whether a microorganism is susceptible, intermediate or resistant to specific antimicrobial medicines. AST results underpin AMR surveillance data.

Antimicrobial use (AMU)

Data describing how antimicrobials are used, including what medicines are used, in what quantities, how often, and by whom. AMU data are commonly collected through surveys such as point prevalence surveys.

Burden of disease (BoD)

A quantitative measure of the health impact of disease, typically expressed using disability-adjusted life years (DALYs), which combine years of life lost due to premature mortality and years lived with disability. In the AMR context, BoD estimates quantify illness and death attributable to resistant infections.

Clinical engagement

Activities designed to strengthen interaction between laboratories, surveillance systems and clinicians or other practitioners, with the aim of improving the interpretation and use of surveillance data in clinical decision-making, treatment guidelines and infection prevention.

Data analysis

The processing and interpretation of surveillance data to identify patterns, trends and implications, including production of outputs such as antibiograms, national surveillance reports and policy briefs.

Data sharing

The dissemination of surveillance data or analytical outputs to relevant stakeholders, including sub-national facilities, national coordination bodies and international reporting platforms.

Data use

The application of surveillance data or analysis to inform decisions, policies, regulations, practices or resource allocation. Data availability alone does not imply data use.

Economic analysis (of AMR)

Analytical approaches that estimate the economic costs, impacts or benefits associated with AMR or AMR-related interventions, including costs to health systems, productivity losses and the economic case for investment in AMR action.

External Quality Assessment (EQA)

An independent process used to assess laboratory performance by testing samples with known results and comparing laboratory outputs against expected standards. EQA participation is a key indicator of laboratory quality.

Facility (as in 'facility level')

A physical site where health or veterinary services are delivered and diagnostic testing is undertaken, such as a hospital laboratory, clinic, veterinary laboratory or equivalent service unit. In this report, facility-level refers to activities, practices or use of data occurring within or directly linked to individual service or laboratory sites, as distinct from national- or system-level processes.

Laboratory strengthening

Improvements to laboratory infrastructure, equipment, quality systems, workforce skills and operating procedures that enable reliable diagnostic testing and surveillance.

National Reference Laboratory (NRL)

A designated laboratory with responsibility for providing technical leadership, quality assurance, training and oversight for diagnostic testing and surveillance within a country.

One Health

An approach that recognises the interconnectedness of human health, animal health and the environment in the emergence, transmission and control of antimicrobial resistance. In this evaluation, One Health is assessed pragmatically, focusing on multisector coordination and surveillance capacity rather than assuming fully integrated systems.

Passive surveillance

AMR data derived from routine clinical or veterinary practice, based on samples taken from patients or animals presenting for care. Passive surveillance was the primary focus of Fleming Fund support in human health.

Point prevalence survey (PPS)

A standardised survey method used to collect data on antimicrobial use at a single point in time within healthcare facilities, commonly used to generate AMU data in human health.

Proficiency testing

A system for objectively checking a laboratory's performance, involving comparison of a laboratory's test results with those of peer laboratories or a reference laboratory, as part of quality management systems.

Results Framework

The structured set of indicators used by the Fleming Fund to monitor progress against agreed outputs and intermediate outcomes across countries, sectors and years.

Surveillance system

The combination of laboratories, data systems, workforce, governance arrangements and reporting processes used to generate, analyse, share and use health-related data on a routine basis.

Sustainability

The likelihood that surveillance systems, capacities and practices supported by the programme will continue after external funding ends, including the ability to finance recurrent costs, retain skilled staff and institutionalise functions.

Throughput

The volume of samples processed by a laboratory over a given period. Throughput affects data quality, workforce proficiency, efficiency and value for money.

Workforce

The cadre of personnel directly involved in laboratory-based AMR surveillance, including laboratory scientists, technicians and associated surveillance, quality and analytical staff; in this report the term is used primarily to refer to people working in or closely supporting laboratories, rather than the wider health or veterinary workforce.

Executive summary

The Fleming Fund is the United Kingdom's flagship investment in strengthening antimicrobial resistance (AMR) surveillance in low- and middle-income countries. Established in 2015 and implemented between 2017 and 2026 (with most activities finishing at the end of 2025), it committed[1] up to £498 million of Official Development Assistance (ODA) to build laboratory-centred surveillance systems capable of generating, analysing and sharing reliable AMR, antimicrobial use and antimicrobial consumption (AMC) data. The Fund operated in over 30 countries across Africa, Asia and the Caribbean, supporting human and animal health and – more recently – environmental and food safety sectors under a One Health ambition. Delivery combined a large portfolio of country, regional and strategic alignment grants, together with a Fellowship scheme, managed by Mott MacDonald on behalf of the UK Department of Health and Social Care (DHSC), and a smaller set of global grants managed directly by DHSC. This report provides an independent summative judgement on what the Fund achieved over 2017–26, why results were realised or constrained, and what this implies for future programming.

Headline judgement

The Fleming Fund has delivered strongly against its core objective of strengthening laboratory-centred AMR surveillance across a large and complex multi-country portfolio. It has significantly increased the volume of AMR data available and strengthened countries' capacity to analyse and interpret that data within both national systems and global surveillance platforms—particularly in human health. In doing so, it has also generated important enabling gains, including strengthened skills, professional networks, standards and technical capabilities that underpin sustained, evidence-informed AMR action. Given the breadth, technical complexity and operating context of the programme, these are substantial achievements.

However, progress beyond these foundational results is more uneven. Contributions to governance and the wider enabling environment are clear and relatively consistent. There are emerging signs of changes in practice and attitudes, but these remain early stage and uneven across settings. By contrast, sustained policy uptake, routine use of surveillance data at national level, and domestic financing for AMR surveillance are limited and variable. Where change has occurred, surveillance data has typically informed decisions alongside other factors, rather than acting as a primary driver.

Looking ahead, prospects for sustaining these gains are constrained and vary widely, particularly where continuation depends on sustained domestic financing for resource-intensive surveillance activities. As a result, overall value for money is mixed, reflecting strong delivery against outputs and foundations, but weaker progress in translating these into durable system-level outcomes.

Performance was strongest where results could realistically be delivered within the programme timeframe and within the Fund's direct sphere of control, such as infrastructure, workforce development and surveillance platforms. It was weaker where results depended on sustained domestic financing, behaviour change or institutional reform. These patterns reflect the Fund's design choices, delivery model and operating context, including compressed implementation timeframes and structural financing constraints, rather than weaknesses in grant-level delivery.

Table 1 summarises the Fund's results and the main factors shaping performance, grouped by core elements of the Fund's Theory of Change (shown in Figure 2). It reflects variation in

[1] Final figures for expenditure were not yet available at time of writing

outcomes achieved, with stronger and more consistent results in outputs, intermediate outcomes and governance-related longer-term outcomes, and more limited and uneven progress in policy, financing and system-level change.

Table 1 Summary results against Fleming Fund Theory of Change areas (use of colours is explained in footnote)[2]

ToC area	Headline judgement	Main explanatory factors (why)
Outputs & foundations (laboratories, workforce, governance)	Targets agreed between DHSC and the Management Agent (MA) were largely met or exceeded. Laboratory strengthening and workforce capacity strengthening were delivered at scale; progress on governance outputs was more variable across countries.	Delivery was strongest where objectives were clearly specified and contractually embedded, reflecting the Fund's primary focus on outputs.
Intermediate outcomes (data generation, analysis, sharing)	Clear progress in the scale and growth of available AMR surveillance data and analysis, particularly in human health, and stronger international data sharing. Other data types beyond AMR were also strengthened.	Intermediate outcomes depended on foundations (labs, quality systems, workforce) and were supported by portfolio features such as Regional Grants providing shared functions and enabling effects (standards, platforms, capacity).
Longer-term outcomes (changes in AMR related policy, practice and funding)	Although the Fleming Fund aspired to support downstream policy and action on AMR, its most consistent contributions lie in governance arrangements and the conditions that make action possible rather than in sustained policy, practice or financing change. Policy and regulatory developments observed during the period were limited, uneven and the Fund was rarely the decisive driver.	This reflects a range of factors including: logical decisions to focus initially on data generation, only later turning to address data use for action on AMR; implementation delays; and the limited extent to which the Fund's core design and delivery model consistently sought to directly influence political priorities, institutional incentives and external drivers that shape AMR action. Expected results for Long-Term Outcomes (LTOs) spread over Phases 2 and 3, but the MA was only contractually responsible for one aspect of these (sustainability).
Sustainability	Prospects are limited and vary substantially, particularly where continuation depends on high recurrent costs and maintenance of activities that the Fund had previously supported.	Structural features made sustainability within a ten-year horizon unlikely; late transition planning, abrupt closure at the end of Phase 2 and a more constrained external funding environment further weakened prospects.
Value for money	Value for money was mixed: stronger on cost control and aspects of delivery efficiency, weaker on effectiveness, equity and sustainability.	Strong cost control supported delivery at scale, but value for money was weaker where benefits depended on sustained domestic financing and complex institutional change, reflecting design and context rather than delivery failure.

[2] The colour scale is used to provide a visual indication on the extent to which results have been delivered, and/or on the prospects for delivery. Dark green = strong results; lighter green = overall progress but with some variation; Orange = some results, but more limited and uneven; Pink = fewer observable results.

Purpose, scope and approach

This summative evaluation assesses Fleming Fund achievements between 2017 and 2026, drawing on earlier evaluative work and new evidence collected during 2025–26. It serves both accountability and learning purposes, assessing performance primarily against the Fund's original intent, design logic and Results Framework, while also recognising additional value that became more visible over time. Judgements are based on triangulation of portfolio-wide monitoring data, in-depth country case studies, and regional and global stakeholder evidence. Methodological constraints, including uneven data quality (e.g. in outcome-level Results Framework indicators) and limited primary data collection during Phase 2, were mitigated through triangulation and a focus on trajectories rather than definitive outcome claims. The analysis also places particular emphasis on explaining observed performance, including the role of programme design, delivery model and operating context in shaping outcomes.

Addressing a broad set of priority evaluation questions within limited space requires that detailed evidence and supporting analysis are set out in the Annexes and wider learning products (see Box 8 in Annex 4.4.5). The treatment in this report is therefore necessarily selective and does not aim to provide a comprehensive account, particularly of outputs and intermediate outcomes. As a result, the breadth of the Fund's achievements is not fully represented here[3], which is an inherent limitation of the report structure. Notwithstanding this, the evaluation recognises substantial and technically strong delivery across implementing partners;[4] however, in line with DHSC's commissioning intent, the analysis focuses on higher-level outcomes, sustainability and value for money.

Context that is critical for interpreting observed results

AMR surveillance is technically and institutionally demanding, requiring functioning laboratories, skilled staff, reliable supplies and governance arrangements that link data to decision-making. These conditions are uneven in many low- and middle-income countries, particularly in animal and environmental health systems operating within a One Health context.[5]

The Fleming Fund's DHSC-led, laboratory-centred design and large, multilayered grant architecture enabled delivery of infrastructure and training at scale, but limited flexibility to engage deeply on policy influence, financing and political incentives. Delivery was further constrained by external shocks, most notably COVID-19, global inflation and supply chain disruption, and by repeated programme replanning and approvals (see Figure 4). As a result, effective implementation time was reduced from ten to around seven years, limiting alignment with the political, institutional and financing horizons typically required to embed national AMR surveillance systems. These features also shaped the extent to which surveillance outputs translated into downstream outcomes, including policy, practice and financing decisions.

The programme also evolved over time. While initial design choices reflected available knowledge and delivery constraints at inception, implementation - particularly during Phase 2 - reflects a process of learning about how to strengthen AMR surveillance systems across diverse country contexts. As one of the first large-scale investments of its kind, the Fleming Fund was necessarily operating in an area with limited prior operational experience. Subsequent strategic shifts (described in Annex 2.4), including increased emphasis on data use, clinical engagement and economic analysis, reflect this adaptive approach rather than a fixed design executed unchanged over time.

---

[3] Further information and resources on the Fund's achievements and learning from its experience are available on The Global Health Network (https://tghn.org/)

[4] Closure reports for all grants can be accessed on devtracker, expected during 2026: https://devtracker.fcdo.gov.uk

[5] See glossary for definition of One Health. More detail is available in Sections 2.2 and 2.3 and Annex 2.5.

Findings

The evaluation's core findings are presented below, assessing what the Fund achieved across its portfolio and why, drawing together evidence on outputs, outcomes, sustainability and value for money. See Section 4 and Annex 4 for details.

Foundations for AMR surveillance. Delivery at output level was substantial across the portfolio. The Fund strengthened laboratory infrastructure, expanded surveillance coverage and delivered large-scale workforce development across human and animal health sectors, and more recently environmental domains. Targets for laboratory and workforce strengthening were largely met or exceeded, with more variable progress observed on governance outputs.

Increased availability of data, analysed and shared. Intermediate outcomes are most evident in the scale and growth of AMR surveillance data, particularly in human health, and in increased international reporting to platforms such as GLASS, ANIMUSE and InFARM. The Fund made a foundational contribution to data flows, standards and shared analytical functions, but increases in data volumes come, as expected, with known caveats (e.g. in terms of selection bias) that influence use at different levels; data are currently more useful at local level than at national and international levels (see Box 2 for more details).

Use of data for action on AMR. Some longer-term outcomes are evident, particularly in strengthened governance arrangements and enabling environments. There is also evidence of changes in practice and attitudes, particularly at facility level, although these remain early and are not yet systematically embedded. Evidence on policy and regulatory change is more limited and uneven. Where such changes occurred, there is illustrative evidence that the Fund contributed to one or more of the drivers of change, often through strengthening technical capacity, professional networks and shared understanding of AMR, rather than through surveillance data acting as a direct or decisive trigger for action.

Sustainability. Prospects for sustaining country-level surveillance activities are limited and vary widely. High recurrent costs, reliance on imported consumables and maintenance, and uneven institutionalisation constrain continuation at scale, particularly in animal health. These constraints reflect both structural features of the Fund's design, including reliance on relatively resource-intensive surveillance models, and wider fiscal and institutional conditions in partner countries. Efforts to maximise sustainability intensified late in Phase 2 but were undermined by compressed timelines, mixed transition signals and the short-notice decision to close the Fund at the end of Phase 2. The most durable legacy is likely to lie in human capital, professional networks, norms and standards that support AMR action through alternative pathways.

Value for money. Value for money[6] is mixed. Cost control and aspects of delivery efficiency were relatively strong across a large and technically complex portfolio at scale. Effectiveness was weaker where results depended on sustained financing, behaviour change and institutional reform, reflecting design choices and operating context rather than grant-level delivery failure. Equity was the weakest VfM dimension, reflecting the absence of an explicit equity lens in Phase 1 and partial progress in Phase 2, with limited integration of equity considerations into core programme design and the distribution of benefits.

[6] Value for Money is assessed using the 4Es framework, considering Economy, Efficiency, Effectiveness, and Equity. *Economy* refers to controlling the cost of inputs; *Efficiency* to converting inputs into outputs without unnecessary friction; *Effectiveness* to the extent outputs led to intended outcomes; and *Equity* to how fairly benefits were distributed.

Conclusion

Overall, the Fleming Fund delivered strongly against what it was designed and contracted to do, particularly in strengthening laboratory capacity, workforce skills and the foundations of AMR surveillance across a large, multi-country and multisector portfolio. Most output-level targets were met or exceeded, and the Fund made significant foundational contributions to global and regional AMR surveillance platforms, standards and professional networks, consistent with its original strategic intent.

Beyond these contracted outputs and foundations, the Fund also generated a set of enabling and more durable effects, including strengthened human capital, professional networks, norms, platforms and governance arrangements, that support AMR action primarily through strengthened surveillance, coordination, evidence sharing and norm-setting over time. These effects are likely to outlast the programme period, even where routine national surveillance systems remain fragile.

Progress on deeper, longer-term system change, such as sustained policy action, routine data use and domestic financing, was limited and uneven. Improved surveillance capacity did not consistently translate into action, reflecting misaligned incentives, weak formal routes from evidence to policy and practice, domestic financing constraints, compressed implementation timelines and external shocks. Sustainability prospects for routine surveillance therefore remain constrained in many settings, particularly where models rely on high recurrent costs.

Taken together, these patterns reflect the Fund's design choices, scale and delivery architecture, and the challenging operating context in which it was implemented, rather than weaknesses in grant-level delivery. Sustainability prospects and value for money vary primarily by intervention type: outcomes that could be delivered and realised within the programme timeframe performed more strongly, while those dependent on longer-term institutional and financing change performed more weakly. Several of the constraints identified in this evaluation were already recognised during Phase 2 and informed early thinking on the direction of potential future Fleming Fund programming.

Lessons

Lessons are intended for designers, commissioners and senior implementers of AMR and other complex global health system programmes. Many are not new, reflecting well-established insights from international development and global health practice, and the experience of the Fleming Fund is not unique within ODA-funded programmes in low- and middle-income countries. Lessons therefore focus on how intent, design and delivery interacted in practice. They highlight how early design and commissioning choices shape what is realistic to achieve within a given funding horizon, rather than pointing to shortcomings in grant-level delivery.

- Fix outcomes early; adapt pathways, not purpose.
- Use grant-making where pathways to achieving agreed results are clear; use different approaches where there's more complexity, uncertainty and need for adaptation.
- Do not assume linear pathways from evidence to action.
- Treat sustainability as a design imperative, not an end-stage activity.
- Calibrate ambition to context and institutional capacity.
- Treat One Health as a design lens, not a delivery template.

1. Introduction

1.1. Antimicrobial resistance and why surveillance matters

Antimicrobial resistance (AMR) is a major and growing global health threat, undermining the effectiveness of routine medical treatment and increasing the risks associated with infection, surgery, childbirth and health system shocks, such as disease outbreaks or sudden service disruptions. It is not just a health issue, but also an issue of economic security, food safety and security. Addressing AMR requires coordinated action across human, animal and environmental health systems, informed by reliable evidence on resistance patterns, antimicrobial use (AMU) and antimicrobial consumption (AMC). However, AMR surveillance systems have been weak, fragmented or under-resourced in many low and middle-income countries, limiting national decision-making and participation in global surveillance efforts. Strengthening AMR surveillance is therefore widely recognised as a necessary foundation for effective national, regional and global action on AMR, even though surveillance alone does not guarantee policy or practice change.

1.2. The Fleming Fund

The Fleming Fund is the United Kingdom's flagship investment in strengthening AMR surveillance in low- and middle-income countries (LMICs). Established in 2015 and implemented between 2017 and 2026 (with most activities finishing at the end of 2025), the Fund committed[7] up to £498 million of Official Development Assistance (ODA) to build laboratory-centred AMR surveillance systems capable of generating, analysing and sharing higher-quality data. The programme operated in over 30[8] countries across Africa, Asia and the Caribbean, supporting human, animal and (more recently) environmental and food safety sectors using a One Health[9] approach. Its primary focus was on laboratory infrastructure, workforce capacity, quality systems, governance arrangements and reporting to global platforms, rather than on delivering downstream AMR control or containment interventions. See Section 2 and Annex 2 for details.

1.3. Purpose and scope of this evaluation

This report presents the independent summative evaluation of the Fleming Fund, assessing what the programme achieved between 2017 and 2026. The evaluation serves two purposes:

1. accountability to Parliament and UK taxpayers for the use of ODA; and
2. learning, to inform future UK and international investments in AMR and other complex global health system programmes.

The evaluation assesses the Fleming Fund primarily against its agreed intent, design logic and stated objectives, as set out in approved programme documentation and subsequent agreed revisions. This provides a transparent basis for accountability, given the programme's scale, duration and technical complexity. In doing so, the evaluation distinguishes clearly between:

- delivery against agreed intent – what the programme was designed and contracted to do;
- additional value that became more visible over time, including enabling effects such as strengthened skills, networks and platforms not fully articulated at programme inception.

[7] Final figures for expenditure were not yet available at time of writing

[8] We refer to 22 countries in subsequent sections, as 22 is the number for which programme monitoring data are available.

[9] See glossary for definition of One Health. More detail is available in Sections 2.2 and 2.3 and Annex 2.5.

This distinction allows the evaluation to maintain accountability to original commitments while also reflecting how the programme's contribution is now understood by key stakeholders, without retrospectively shifting the basis of assessment.

The evaluation focuses primarily on country-level results, while also considering relevant regional and global influences where these plausibly shaped country outcomes. Within this scope, the evaluation addresses the following six core questions.[i] More details on the evaluation methodology are found in Section 3 and Annex 3.

- EQ1: To what extent has there been an increase in the generation, analysis and sharing of quality AMR-related data at country level, and what has been the Fleming Fund's contribution?
- EQ2: To what extent have the Fund's investments been aligned with national priorities and coherent across the portfolio and with other relevant investments?
- EQ3: How likely are the Fleming Fund's country-level results to be sustained, and why?
- EQ4: To what extent has improved availability of AMR-related data influenced policy, practice, attitudes or financing decisions?
- EQ5: What has been the increase in quality AMR-related data shared and reported internationally, and why?
- EQ6: Did the Fleming Fund's country-level investments offer value for money?

1.4. Structure of the report

The remainder of the report is structured as follows:

- Section 2 (Background and context) sets out the programme's design choices, assumptions and the external factors shaping implementation.
- Section 3 (Methodology) describes the evaluation approach, analytical framework, evidence base and limitations in detail.
- Section 4 (Findings) presents the core evidence on outputs and foundations, effectiveness at intermediate and longer-term outcome levels, sustainability and value for money.
- Section 5 (Conclusions) draws together judgements on performance, contribution and limitations.
- Section 6 (Lessons) distils transferable learning for designers and commissioners of future AMR and other complex global health system programmes.

Additional technical detail and supporting evidence for the findings presented in this report are provided in Volume II, Annexes.

Additional insights and learning from the Fund's ten years of experience is set out in a range of learning products available at The Global Health Network (TGHN).[ii] A journal series entitled 'Antimicrobial Resistance (AMR) surveillance: Ten Years On' will also be published, subject to peer review, in Clinical Infections Diseases (CID)[iii] by the end of 2026.

2. Background and context

This section provides contextual detail to interpret the findings that follow, including core design choices, assumptions and external factors that shaped what the Fund was positioned to deliver.

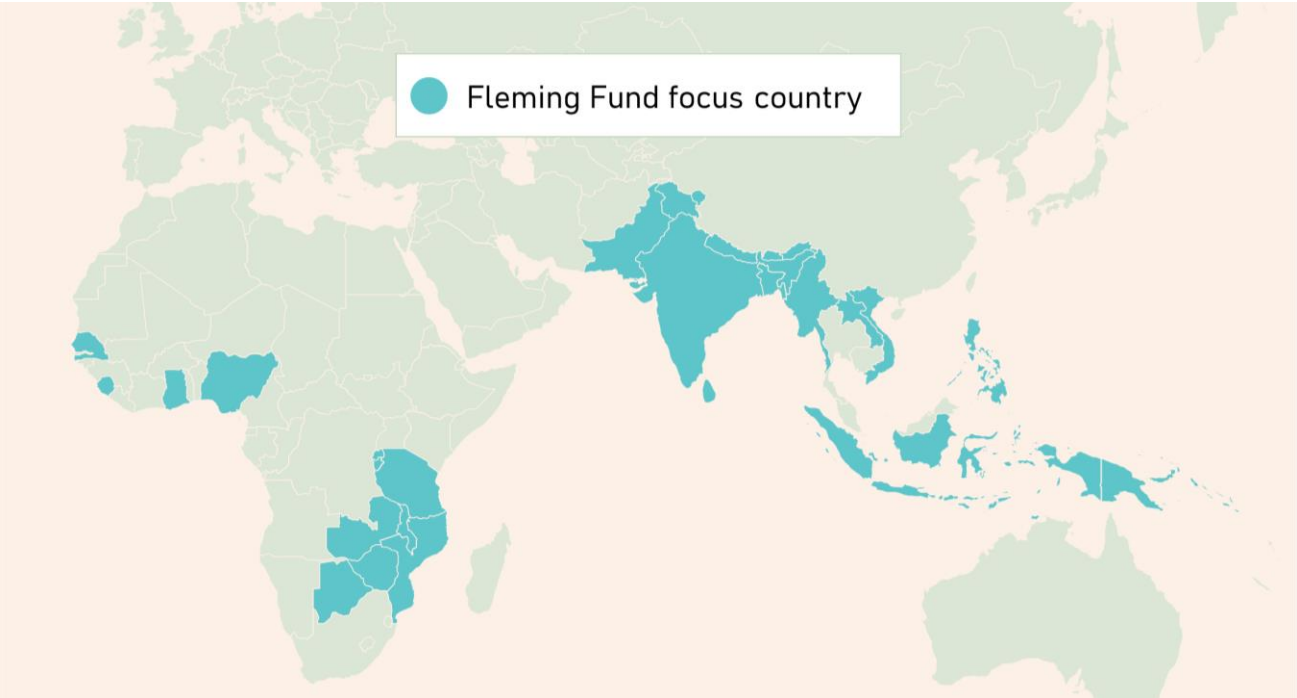
2.1. AMR surveillance as a global and national challenge

Surveillance is a necessary but demanding component of an effective AMR response,[iv] requiring functional laboratories, skilled personnel, reliable supply chains, quality assurance systems and governance arrangements that link laboratory outputs to decision-making. Across many LMICs and sectors, these capabilities were historically uneven or under-resourced, and existing AMR surveillance systems, where they existed, were often fragmented, incomplete or not routinely used to inform policy or practice.[v] Strengthening surveillance systems was therefore widely understood as a prerequisite for more effective national and global action on AMR.

2.2. The Fleming Fund: purpose and scope

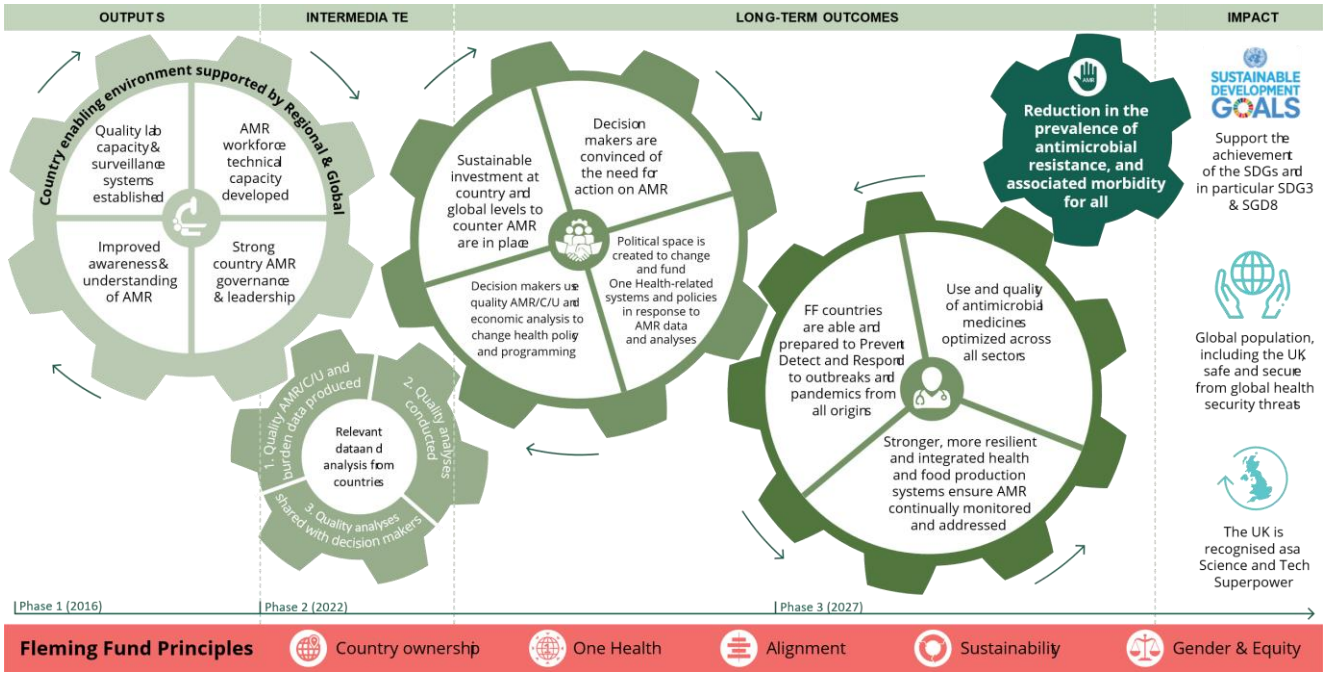
In response to recognised gaps in AMR surveillance capacity, the UK Government established the Fleming Fund in 2015 as its flagship investment in AMR surveillance in LMICs (Figure 1). Funded through ODA and managed by the Department of Health and Social Care (DHSC), the Fund was designed primarily as a laboratory and surveillance-system-strengthening programme, rather than a comprehensive AMR control intervention.

Figure 1 Fleming Fund focus countries



Across two phases of implementation (2016–2026), the UK committed up to £498 million – making the Fleming Fund the largest national investment in AMR surveillance globally – to enable supported countries to generate, analyse, share and ultimately use higher-quality AMR data through national systems aligned with international standards. Its scope expanded over time, including through a set of strategic shifts at the start of Phase 2 (see Annex 2.4 for details) to include greater attention to antimicrobial use and consumption data,[vi] burden of AMR and economic analyses, and data use for policy and practice (see Figure 2).

Figure 2 The Fleming Fund's portfolio-level Theory of Change



2.3. Core design choices and assumptions

A set of foundational design choices shaped what the Fleming Fund was structurally positioned to deliver. See Annex 2.3–2.8 for details.

First, the programme adopted a DHSC-led, laboratory-centred model, reflecting both the nature of the identified evidence gap in AMR surveillance and areas where the UK had well-established technical and institutional capability. This approach provided clarity and measurability for a large, multi-country ODA investment, with a primary focus on laboratory infrastructure, equipment, workforce training and quality systems, concentrating efforts on technical foundations rather than on surveillance data use in policy and practice.

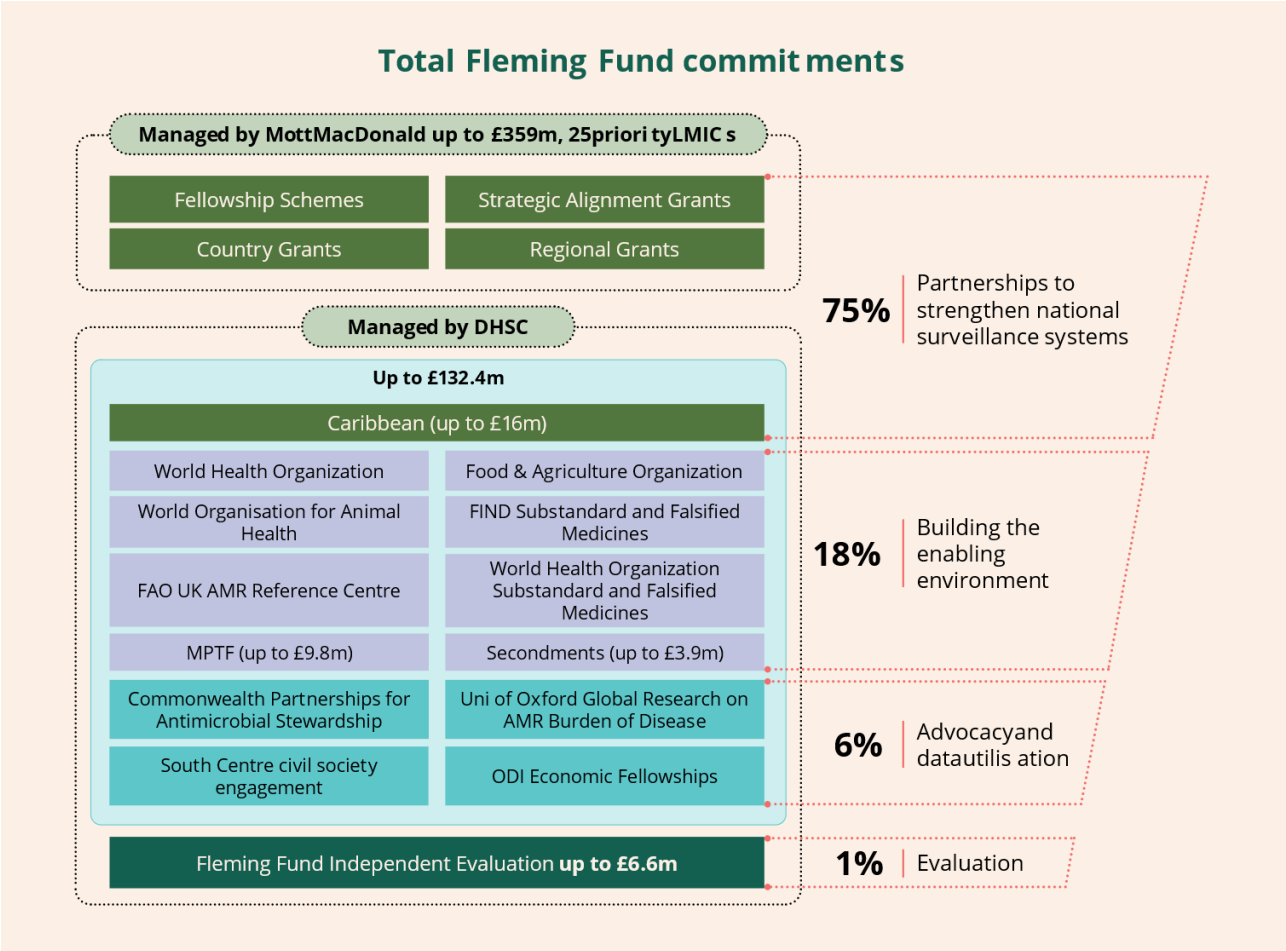
Second, delivery was organised through a large, multilayered grant architecture, combining country-, regional-, strategic alignment- and direct grants (Figure 3) across a set of ODA-eligible countries in Africa and Asia. The programme adopted a defined technical scope: supporting detection and reporting of a prioritised set of pathogen–antimicrobial combinations aligned with World Health Organization (WHO) surveillance standards. It funded a standardised set of activities focused on strengthening laboratory and surveillance functions within existing national public health systems. Implementation was channelled primarily through competitive grant-making managed by a contracted Management Agent (MA), Mott MacDonald, and a set of grants, managed directly by DHSC, to international organisations; there was no direct financial

disbursement to governments.[10] This architecture enabled technical reach across countries and sectors but generated significant coordination demands and extended planning, contracting and approval chains.

Third, contracting and accountability arrangements were strongest at output level, with clear targets agreed for laboratory strengthening, workforce development and related deliverables within this delivery model. Broader outcome ambitions, such as routine use of surveillance data to inform policy, practice or financing decisions, were articulated more fully as the programme evolved. However, this occurred after major delivery and contracting arrangements were already in place, reflecting emerging learning during implementation and the programme's adaptive understanding of effective AMR surveillance. This created a clear distinction between what was contractually specified and managed through the grants, and wider system-level expectations that relied on institutional change, cross-sector coordination and domestic decision-making beyond implementers' direct control.

As implementation experience accumulated, clearer operational understanding emerged regarding how to strengthen AMR surveillance systems across different sectors and contexts. Several key interventions and areas of emphasis were therefore developed or scaled relatively late in the programme, limiting the time available for their effects to be observed.

Figure 3 Fleming Fund grant portfolio (source: DHSC financial reporting)



Programme-level assumptions about how outcomes and impact would be achieved were not documented at the outset, either in a business case or theory of change (ToC). A programme-

[10] Except to the Royal Government of Bhutan; also note some public-sector institutions were sub-grantees (e.g. national public health institutions (NPHIs) and universities).

level ToC and narrative were developed later, including several implicit assumptions that shaped the programme's trajectory and risk profile.

- Increased availability of surveillance data would naturally stimulate action on AMR.
- A laboratory-centred model should be the entry point across diverse LMIC contexts.
- Strengthened technical capacity gains would be sustained through a focus at the individual level (rather than organisational or institutional levels), without explicit planning from the outset.

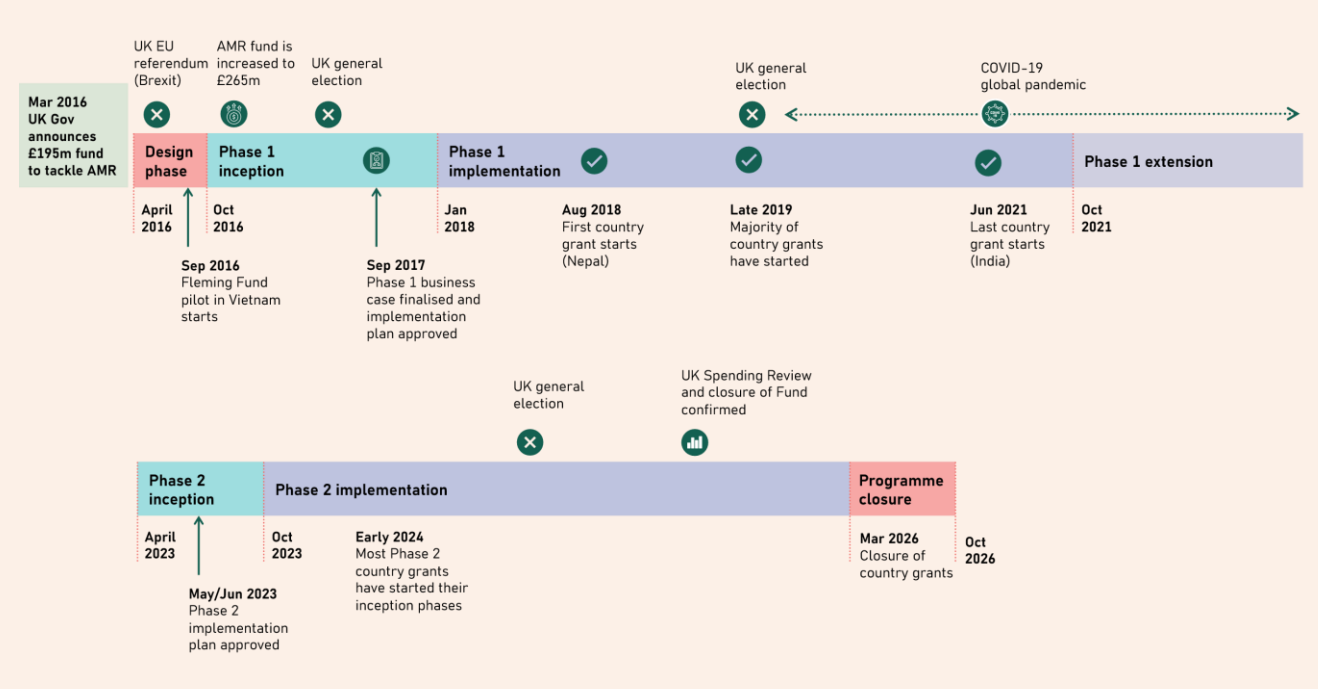
Early programme design also reflected prevailing scientific assumptions at the time regarding the scale and significance of zoonotic transmission of AMR, and the role of animal surveillance in mitigating public-health risk. On this basis, progress across sectors was expected to proceed in parallel under a One Health framing, despite differences in system maturity, incentives and financing. While the Fund articulated a One Health ambition from the outset, early design assumptions reasonably but imperfectly treated human and animal health systems as more symmetrical than proved to be the case. In practice, the two sectors differ markedly in institutional mandates, financing arrangements, professional incentives and the value to stakeholders of AMR surveillance outputs. Human health AMR surveillance typically aligns with established public-health functions, clinical decision-making and international reporting incentives, whereas animal health AMR surveillance often relies on weaker mandates, more fragmented delivery structures and fewer direct links to routine decision-making.

2.4. External factors shaping implementation

The Fleming Fund operated through a period of exceptional volatility that materially affected implementation, reflected in Figure 4. The COVID-19 pandemic disrupted laboratory operations, diverted capacity to emergency response and delayed planned activities, while global inflation and supply-chain pressures increased costs and lead times. In parallel, domestic political changes in the UK resulted in repeated cycles of spending review, replanning and approval, contributing to uncertainty and delay.

Although nominally a ten-year programme, the combined effect of design timelines, reprogramming, approval cycles and external shocks meant that effective implementation time was substantially shorter, and in some countries significantly so (see Annex 2.7 for details). This compressed timeframe was misaligned with the political, institutional and financing horizons required to embed complex system reforms, particularly those emphasised in Phase 2.

Figure 4 Implementation timeline



2.5. Implications for delivery

The Fleming Fund's delivery architecture had predictable implications for overheads and transaction costs (see Section 4.5). The programme was designed around a single, competitively procured MA responsible for grant-making, coordination and assurance across a large multi-country portfolio. At the time of design, MA delivery models were in use across other comparable ODA funds. Independent Commission for Aid Impact (ICAI) reviews[vii] of these funds treated such models as appropriate mechanisms for managing fiduciary risk, consistency and economy at scale, focusing scrutiny instead on strategic clarity, results management and value-for-money arrangements. However, evidence across ODA funds such as CSSF and the Prosperity Fund indicates that complexity and diffuse accountability are common to many multi-partner programmes, regardless of whether formal MAs are used.

However, operating this model through DHSC oversight and systems - characterised by strict year-by-year budgeting constraints and extensive assurance and compliance requirements - necessarily increased coordination, due diligence and grant-management costs. Maintaining readiness across countries, sequencing grants to manage risk, and responding to slippage within fixed financial years required a sizeable coordination and control infrastructure. These features represent conscious design trade-offs shaped by system-level and contextual constraints, rather than delivery inefficiency or weak programme management. In addition, these challenges wouldn't be considered exceptional in the case of the Fleming Fund in comparison to other ODA programmes. See Annex 2.3 for detail.

Beyond their implications for overheads and transaction costs, the Fund's design choices and operating context also established the parameters within which results were achievable. The laboratory-centred delivery model enabled substantial delivery of infrastructure, training and foundational capacity, but limited the depth of engagement on data use, policy influence and long-term financing. In combination with variations in baseline capacity and country ownership, modest country-level allocations relative to ambition, later expansion of scope in Phase 2, and repeated replanning in response to external shocks (as discussed in section 2.4), these features

shaped uneven trajectories across countries and sectors and constrained opportunities to consolidate the priorities reflected in the strategic shifts (see Annex 2.4 for details).

3. Methodology

3.1. Purpose and scope of the evaluation

This summative evaluation assesses what the Fleming Fund achieved between 2017 and 2026, spanning both phases of implementation. It builds on earlier evaluative work, including the Phase 1 Summative Evaluation[viii] and the 2024 Progress Report,[ix] and incorporates additional evidence collected during 2025–26.

The evaluation addresses both accountability and learning objectives. It focuses primarily on country-level results, while also considering regional and global interventions where these plausibly influenced country outcomes.

Performance is assessed against the Fleming Fund's stated objectives,[x] design parameters and Results Framework, as set out in approved programme documentation and agreed revisions over time. This provides a consistent and transparent basis for accountability given the scale, duration and technical complexity of the programme.

3.2. Analytical framework and evaluation questions

The evaluation questions listed in Section 1 are mapped to an AMR surveillance results chain (see the Fund's Theory of Change in Annex 2.1), distinguishing between:

- **Outputs and foundations:** strengthening laboratories, workforce capacity, governance arrangements and data systems;
- **Intermediate outcomes:** improvements in data generation, quality, analysis and sharing at facility, national and international levels;
- **Longer-term outcomes**: examining whether improved data availability contributed to changes in policy, practice, attitudes or financing decisions (expected over Phase 2 and Phase 3); and
- **Sustainability and value for money**:[11] assessing the likelihood that results will endure and whether resources were used effectively.

This framework helps clarify what can reasonably be expected within the programme timeframe and supports differentiated judgements across outputs, outcomes and sustainability.

3.3. Evidence base and data collection

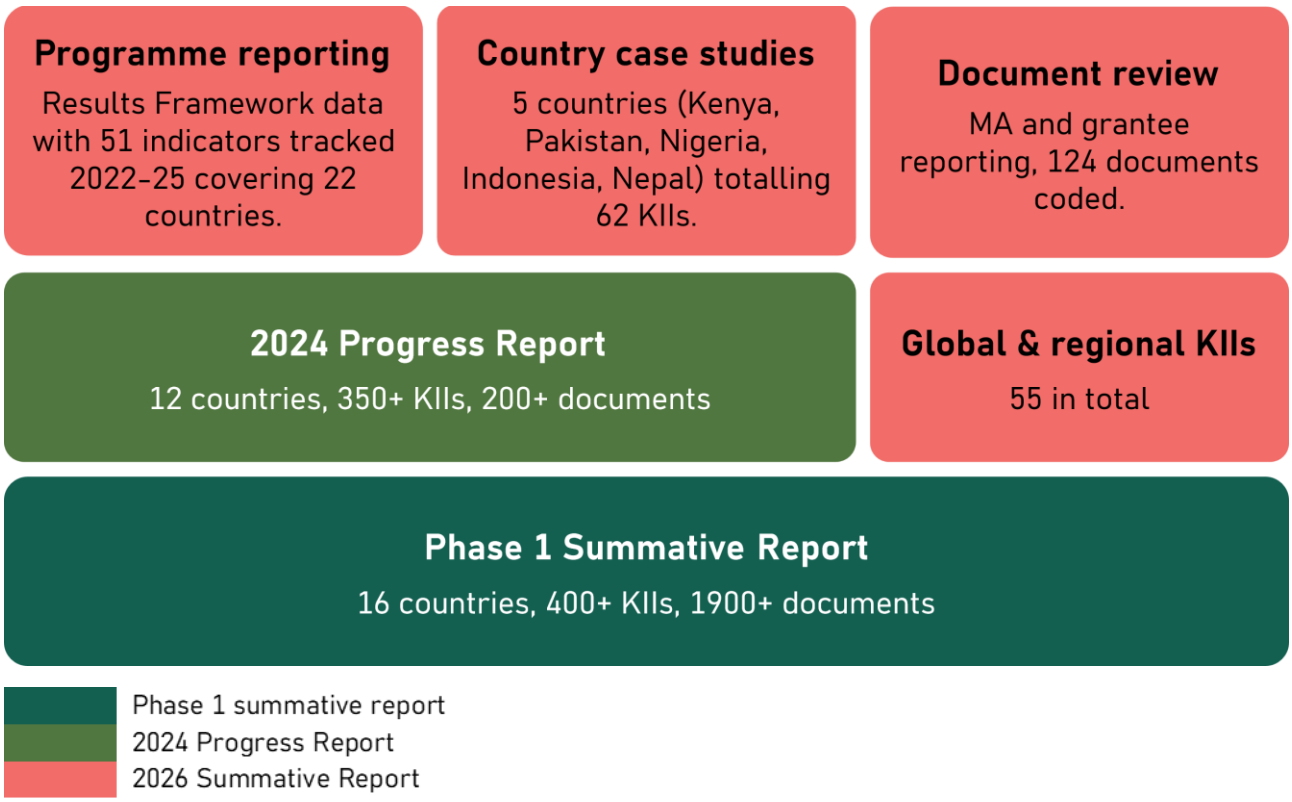
The evaluation adopted a mixed-methods approach, drawing on multiple sources of evidence:

- structured review of programme documentation, including grant designs, monitoring reports and strategic documents;
- Results Framework data covering all Fleming Fund countries;
- in-depth country case studies combining document review and interviews with national stakeholders; and
- global and regional key informant interviews with multilateral organisations, technical partners, the MA and DHSC.

[11] See footnote 2 in Executive Summary.

Five country case studies were selected to reflect diversity in geography, baseline surveillance capacity and implementation trajectories; this is in addition to 12 in-depth country studies completed for the 2024 Progress Report and 16 for the Phase 1 summative report. Global and regional interviews were used to contextualise country-level findings within wider AMR surveillance and governance systems. A summary of the available evidence base is included in Figure 5; additional detail on country samples and document review is included in Annex 3.4.

Figure 5 Summary of evidence base for this report



3.4. Analysis and triangulation

The evaluation applied a range of analysis approaches including contribution-focused analysis (to examine how and where the Fund plausibly contributed to changes in policy and regulation, EQ4a), plus thematic analysis across all EQs, and VfM analysis for EQ6. Cross-case synthesis was used to identify common patterns and contextual variation across countries and sectors.

Findings were triangulated across data sources to strengthen credibility and reduce reliance on any single perspective. Where outcome-level evidence was incomplete or emerging, the analysis focused on trajectories and patterns, rather than definitive impact claims.

Some numerical aggregates presented in this report differ modestly from MA reporting products due to differences in data cut-off dates, revisions during quality assurance, treatment of outliers, and aggregation methods (e.g. site-level vs country-level compilation), for example in Section 4.2. These differences do not materially affect the evaluation's findings or conclusions, which are based on patterns, trajectories and triangulated evidence rather than on point estimates. Where quantitative indicators and qualitative evidence point in different directions, greater evidential weight has been placed on triangulated qualitative evidence to interpret what indicators likely represent in practice.

The evaluation faced several limitations, which are important for interpreting the findings.

- Primary data collection for this Phase 2 summative report was constrained by time, budget and feasibility, resulting in greater reliance on remote interviews and document review.
- Many Phase 2 priorities, particularly those relating to strategic shifts, data use outcomes and sustainability, were at an early stage of implementation at the time of data collection, despite this being close to the end of the programme, limiting the availability of mature outcome-level evidence.
- The evaluation focuses on programme performance and system strengthening rather than microbiological performance at laboratory level; as a result, many findings draw on proxy indicators (such as EQA participation, SOP availability, training, and data volumes), which provide evidence of system development but do not directly measure laboratory diagnostic performance.
- Data quality and completeness varied across countries, sectors and indicators. While the Results Framework is generally available for the whole portfolio, qualitative analysis generated for this stage of the evaluation is available in fewer countries (see Figure 5 for details). In addition, reporting was available only up to the end of 2025, while implementation continued into the first quarter of 2026.
- Addressing a broad set of priority evaluation questions within limited space requires that detailed evidence and supporting analysis are set out in the Annexes and wider learning products (see Box 8 in Annex 4.4.5). The treatment in this report is therefore necessarily selective and does not aim to provide a comprehensive account – particularly of outputs and intermediate outcomes. As a result, the breadth of the Fund's achievements is not fully represented here, which is an inherent limitation of the report structure. Notwithstanding this, the evaluation recognises substantial and technically strong delivery across implementing partners; however, in line with DHSC's commissioning intent, the analysis focuses on higher-level outcomes, sustainability and value for money.

These limitations were mitigated through triangulation across multiple sources, albeit with light referencing and discussion of evidence quality in the main report and reliance on the Annexes. Where results data was incomplete, the analysis focused on trajectories rather than definitive impact claims. See Annex 3.5 for more details. Additional information on the evaluation methodology is also available in the Phase 2 evaluation inception report (Itad, 2024).

4. Findings

This section provides an assessment of progress in strengthening generation of quality AMR surveillance data, analysis and sharing (Section 4.2) and use of that data for policy, practice, financing and governance (Section 4.3). This feeds into assessments of prospects for sustainability (Section 4.4) and value for money (Section 4.5). Findings reflect what is observable from available programme-level monitoring data, which may not capture all of the Fund's activity.

4.1 Progress against agreed outputs and targets

Across the portfolio, through MA- and DHSC-managed grants, the Fund supported strengthening of human health, animal health and environmental surveillance capacity in 22 countries. Activities supported the strengthening of 269[12] laboratories, establishing functioning multisector governance arrangements in most countries (17/22), and delivering more than 80,000 trainings to relevant staff (in addition to more than 375 Fellows supported) to set up and manage national AMR surveillance systems. Outputs have also been delivered at regional and global levels, including to strengthen technical training provision, establish whole-genome sequencing capacity, provide EQA services, and establish global data platforms and normative guidance to support country-level action.

4.1.1 Country-level outputs

Laboratory strengthening. Targets related to strengthening laboratories were generally met or exceeded. Progress was most consistent in human health laboratories, where baseline systems were stronger and surveillance activity could be scaled more rapidly. In animal health laboratories, delivery was also substantial but more uneven, reflecting weaker baseline systems, fewer supported sites, and greater reliance on resource-intensive active surveillance models.

Workforce development. Workforce development targets were consistently exceeded. Large numbers of laboratory, surveillance and related personnel (including policymakers) were trained through a combination of short courses, mentoring, and Fellowship-based approaches. This contributed to improved technical capability and the formation of professional networks within and across countries. While training outputs were strong, the extent to which these gains will be sustained depends on continued system functionality and resourcing (see Section 4.4).

Governance arrangements. Progress in establishing or strengthening national AMR governance and coordination mechanisms, including establishing AMR National Action Plans (NAPs),[xi] was more variable. In many countries, multisectoral coordination bodies were established or made more functional. In others, progress was slower or uneven, reflecting differences in political prioritisation, institutional mandates and cross-sector coordination capacity.

Overall, output-level delivery compares favourably with targets agreed between DHSC and the MA. Where targets were not fully met, shortfalls were generally concentrated in the final year of Phase 2 and related to compressed implementation time, external disruption and late-stage shifts in programme emphasis. For more detail on delivery of outputs, target setting and performance against targets, see Figure 5 supporting narrative in Annex 4.1.

[12] This is the number of HH and AH sites that reported having progressed through or maintained and accepted level of competence (drawn from Results Framework reporting). It is not the total number of sites supported which was 335 in total: 215 HH sites; 94 AH sites; 26 other sites (10 environment; 9 food safety; 7 aqua).

Box 1 Fleming Fund in Nigeria

Fleming Fund (FF) support in Nigeria was closely aligned with NAP 2.0, with activities directly linked to its implementation plan. This strengthened national ownership and ensured AMR and One Health efforts were embedded within statutory systems and partner coordination mechanisms.

Key achievements include:

- **Improved data management:** Strengthened processes and increased confidence in data sharing, particularly in the animal health sector where this was previously limited. Embedding data management roles within existing laboratory functions enhanced sustainability beyond the grant.
- **Enhanced national data flows:** Routine reporting from laboratories to national level improved, although visibility on global platforms remained limited.
- **Increased technical capacity:** High levels of confidence in AMR surveillance methods (e.g. PPS, BoD) across stakeholders, including senior laboratory leadership.
- **Catalysed local sustainability:** Capacity gains are driving proactive efforts to secure resources, including research collaborations and facility-level financing mechanisms such as reagent revolving funds.

4.1.2 Regional- and global-level outputs

Alongside and, in some cases, contributing to country-level delivery, a substantial share of Fleming Fund activity and outputs was delivered through regional, strategic alignment and direct grants operating at regional and global levels. These instruments were intended to provide shared technical functions and public goods that could not be delivered efficiently through individual country grants alone, and to strengthen the wider AMR surveillance ecosystem within which national systems operate.

Regional-level outputs focused on strengthening core technical functions that support country surveillance systems. These included provision of regional EQA services, development and dissemination of standardised laboratory protocols and quality systems, delivery of specialist and advanced training (including whole-genome-sequencing–related capacity), and establishment of regional technical networks and communities of practice. Regional initiatives also supported analytical capability and practical tools to help countries convert multisectoral surveillance outputs into usable evidence, complementing country-level investments in laboratories and workforce.

Global-level outputs, delivered primarily through direct grants managed by DHSC, supported the development, strengthening and operation of global surveillance and data-sharing platforms, alongside production of normative guidance, standards and analytical methods, and support to international coordination and knowledge exchange. Direct Grants were deliberately heterogeneous, reflecting the need to work through multilateral organisations and specialist institutions with appropriate mandates and technical reach. As a result, outputs were oriented primarily towards strengthening global systems, standards, platforms and analytical capacity, rather than delivering uniform country-level results.

Taken together, regional- and global-level grants delivered a wide range of tangible outputs – services, tools, standards, training provision, analytical capability and shared platforms – designed to complement and amplify country-level delivery, reduce duplication, and strengthen the consistency, credibility and interoperability of AMR surveillance across countries and sectors. Unlike country-level grants, output targets were defined at programme level (in the

Results Framework) for regional and global grants, reflecting their role in delivering shared functions and public goods rather than uniform, countable country-level outputs. Further detail on the scope and delivery of regional and global grants is provided in Annex 2.2.

Interpreting output-level delivery

Much of the delivery narrative in this section draws on MA-managed country grants, reflecting their greater visibility through Results Framework data and standardised reporting. Evidence on Direct and Regional Grants is less readily captured through uniform indicators.[xii] These contributions are nonetheless integral to the Fund's delivery model and are summarised separately, with supporting evidence presented in Annex 2.2.

The pattern of output-level delivery reflects both effective execution and the alignment between agreed objectives and the Fund's design. Output objectives were clearly specified, contractually embedded and supported by robust grant management systems, and could be delivered within the available timeframe with limited dependence on domestic political and financing conditions. Where delivery was weaker or incomplete, this was concentrated in areas affected by compressed implementation time, external disruption and late-stage shifts in programme emphasis, as discussed in Section 2.4. These features are important for interpreting subsequent findings on outcomes, sustainability and value for money, which address areas of change that were structurally more complex and less directly within the programme's control.

Output-level delivery under the Fleming Fund compares favourably with that reported for several other cross-government UK aid funds of a similar scale and complexity. This is in contrast to funds such as the Newton Fund[xiii] and the Prosperity Fund,[xiv] where ICAI[xv] identified weaknesses in design alignment, target-setting and portfolio-level results management. The Fleming Fund's output objectives were clearly specified, embedded within grant and MA contracting arrangements, and actively managed, enabling more consistent monitoring and delivery of agreed outputs within the available timeframe.

4.2 Effectiveness: Intermediate outcomes

The strong performance observed at output level (Section 4.1) provides the foundation for assessing progress in the Fleming Fund's intermediate outcomes, focusing on improvements in the generation, quality, analysis and sharing at facility, national and international levels of different types of AMR-related data. It distinguishes between AMR data (patterns of resistance identified through laboratory testing), antimicrobial use (AMU) data describing how antimicrobials are used in practice, and antimicrobial consumption (AMC) data capturing aggregate volumes of antimicrobials sold, supplied or consumed, typically at national or sector level; together these inform national and international AMR surveillance.

The findings in this section reflect what could be observed through available programme-level monitoring data, grant reporting and stakeholder interviews. These sources provide a robust basis for assessing trends and patterns in data generation, quality, analysis and sharing, but they do not capture all Fund activity or all uses of AMR-related data across countries and institutions.

Observed variation in the availability and use of AMR-related data reflects more than differences in technical capacity. Across the portfolio, whether microbiology or surveillance data were generated, analysed and used depended on political prioritisation, financing decisions and demand-side incentives, which varied by context and over time.

Intermediate outcomes improved in areas directly influenced by the Fund, most notably in the scale of AMR data generation and international reporting. Data sharing from surveillance sites to national coordination mechanisms increased, but data was not always

representative or in a form that governments could readily use to inform AMR related surveillance review, planning, prioritisation or regulatory discussion. Sharing of analysed data within facilities in human health was strong in many countries, informing local decision-making and improving practices, but it was weaker in animal health. These results broadly meet agreed intermediate-outcome expectations where indicators were specified and reflect a material contribution by the Fund.

4.2.1 Progress in data generation

AMR data generation – substantial increases achieved, particularly in human health

AMR data generation scaled up substantially, particularly in human health, but gains were uneven across sectors and do not on their own indicate surveillance systems that are nationally representative or fit for routine decision-making. As shown below, AMR data generation at Fleming Fund-supported sites increased substantially over the programme period, to varying degrees by sector. See Table 2 and Annex 4.2.1 for details.

- In **human health**, the scale of growth is clear: across 22 countries, annual samples rose from 149,000 in 2020 to 2.138 m in 2025 (more than a 13-fold increase).
- In **animal health**, total sample numbers were lower (as expected given animal health baseline capacity and smaller numbers of animal health laboratories supported) and trends less consistent; but there was still a six-fold increase in animal health samples (active and passive combined). This reflects the programme's early emphasis on active surveillance, which is conducted in rounds and is inherently resource-intensive.
- **Environmental and food safety** surveillance began in 2024 and expanded rapidly. Reported environmental samples reached more than 18,000 samples, with volumes steady in 2025; food safety sampling reached approximately 53,000–55,000 samples, more than a three-and-a-half-fold increase in one year. These remain smaller than human health AMR volumes but represent meaningful early expansion in new domains.
- **Average sample throughput per laboratory per day suggests a more cautious picture of total sample volumes, and was highly variable across laboratories and countries.** Throughput was much higher in human health but varied from an average of two (in Ghana) to 106 (in Pakistan) samples per lab per day (average of 33) across all Fund countries in 2025. Animal health throughput fluctuated but remained limited in most countries, with a maximum of ten samples per lab per day (Kenya) and an average of three samples per lab per day across all countries in 2025. In both sectors many laboratories seem to be operating below their potential capacity, even when considering the context (see Vol 2, Annex 4.2.1 and Box 7 for discussion on drivers of throughput).

Across sectors, the evaluation found that increased data generation reflects substantial progress in surveillance production capacity (described in Section 4.1). At the same time, key stakeholders note that data volumes alone are not a reliable proxy for surveillance usefulness at national level, given constraints on representativeness and integration (see Box 1 for more details).

Table 2 Samples from Fleming Fund-supported sites by sector 2020–2025

	2020	2021	2022	2023	2024	2025	Trends Over Time
HH	149,468	785,597	896,646	1,137,427	1,828,123	2,138,126	
AH Active	3,281	20,839	29,564	13,194	20,302	36,451	
AH Passive	8,020	26,395	30,001	41,761	38,422	34,770	
EH	0	0	0	0	5,914	12,542	
Food Safety	0	0	0	0	11,901	41,878	
Sample Generation Comparison across Sectors 2020-2025							

AMU and AMC data generation – partial progress, limited routinisation

Generation of AMU and AMC data became a focus later in programme implementation, reflecting increased emphasis in Phase 2 and the absence of routinised national systems for regular data production. As a result, progress in this area was more limited. More AMU and AMC data are available now than earlier in the programme, though progress is generally more limited than for AMR. This pattern is consistent with the Fund's design, since AMU/AMC became a stronger focus only in Phase 2. For **human health** AMU, point prevalence surveys (PPS) increased in coverage across countries by 2025; for **animal health** AMU, more countries reported conducting farm-based AMU surveillance by 2025. For **AMC**, routine national-level data remain limited in many settings, and where AMC exists it is often based primarily on imports rather than comprehensive consumption. See Annex 4.2.1 for details.

Burden of disease and economic data generation – enabling work progressed, but routine national use remains limited

Progress in generating burden of disease and economic analysis was largely in enabling work rather than production of actual analyses, with limited evidence that countries were using these analyses routinely during the delivery window. Progress in routine burden-of-disease and economic evidence is more limited and remains at an earlier stage of maturity, reflecting the late Phase 2 emphasis and the enabling nature of much of the work. For **burden of disease**,[13] adoption of protocols increased over Phase 2. This does not itself indicate routine national or local production of burden estimates, but the Fund did support global modelling work through the GRAM direct grant, which contributed to global burden estimates (see Annex 2.2 for more details). For **economic analysis**, progress is less visible in routine reporting and has largely taken the form of enabling work (methods, tools) with limited evidence of routine national uptake. However the Fund also supported important economic analysis through the EcoAMR series.[xvi]

[13] It is important to note that reference to burden of disease work here refers to country-level work and not to GRAM modelling also funded by the Fleming Fund. See Annex 2.2 for more detail on GRAM.

Box 2 Notes on interpretation and use of generated data[xvii]

The programme generated substantial volumes of surveillance data and strengthened laboratory systems, enabling practical facility-level use in some settings. However, key informants reiterated known limitations in terms of what the outputs from supported AMR surveillance systems can be used for, at this stage.

AMR data drawn from the sentinel surveillance model supported by the Fund are subject to selection bias because testing is not performed uniformly across all eligible patients. Observed resistance patterns reflect the subset of patients who underwent testing, which is influenced by many factors, including operational constraints, clinician test-ordering practices, and care-seeking behaviours. These factors introduce selection bias that may affect both the observed prevalence and the distribution of antimicrobial resistance, and the extent of caveats required or limitations observed in using data at different levels (generally current datasets are more easily used at facility-level than national-level). These factors were known limitations from the outset; . Alternative approaches that mitigated these limitations would have had substantial resource and sustainability implications, and/or required substantially longer Fleming Fund implementation to implement.

Access to usable datasets remains a constraint: international stakeholders reported difficulty establishing what data exist, accessing up-to-date datasets, and producing consolidated national-level profiles. Throughput also varies widely, with implications for efficiency and sustainability: low throughput can signal laboratories operating below capacity and raises concerns about the affordability, sustainability and justification of high-cost automated equipment, human resources, and systems. Overall, the Fund contributed to progress towards surveillance capability rather than a fully policy-ready surveillance system.

4.2.2 Quality of AMR data generated – improving but still uneven, particularly outside human health

AMR data quality improved over time but remains uneven across countries and sectors, with more consistent progress in human health than in animal health. Many AMR programmes focus principally on data volume and less on data quality, and the Fleming Fund consideration of quality is notable. Overall, systems to ensure the quality of AMR data have strengthened over time, and available evidence[14] indicates improvements in data quality and reliability. High-quality data are essential for credible AMR surveillance and downstream use. At national reference laboratory (NRL) level, participation in **EQA** increased in both sectors; accreditation increased between 2024 and 2025 but was not achieved in all NRLs. At site level, more countries participated in EQA schemes and more sites achieved acceptable performance. Evidence on standard operating procedure (**SOP**) availability and training shows mixed but generally improving practice, with gaps existing between having SOPs in place and training staff to apply them consistently. Underpinning evidence, including sector disaggregated data, is presented in Annex 4.2.2.[xviii]

Sustained quality depends on systems and protocols, including before samples reach the laboratory, plus reliable infrastructure, consumables, sufficient throughput, and staff retention.

[14] Note that our analysis of improvements in quality, e.g. on EQA, uses participation in EQA processes and the accreditation status of participating laboratories as proxies for quality. Limitations associated with his are discussed in section 3.5. A more detailed assessment of quality systems, which could for example have usefully considered AST methodology standardisation, CLSI versus EUCAST implementation, internal quality control performance, evaluation of measurement uncertainty, organism identification accuracy, clinical interpretation of AST results, was not possible within time or resources.

The Fund contributed strongly to infrastructure and quality systems, but, unsurprisingly, results varied according to differences in wider system capacity.

4.2.3 Progress in data analysis

Analytical output increased, but depth, integration and routine interpretation remain constrained by capacity, staffing and time pressures. Analysis and reporting of surveillance data increased over time, consistent with a stronger emphasis on analysis in Phase 2. Progress varies by country and sector, and analysis is generally more developed for AMR than for AMC and AMU (see Annex 4.2.3 for details). By 2025, many countries reported producing national reports and briefs containing analysis in human health and animal health, with fewer reporting One Health analyses. In several settings, surveillance outputs supported facility-level analyses such as antibiograms, and some countries developed dashboards and public-facing analytical tools. At the same time, analysis of AMC and AMU data remains weaker and more uneven, and analysis in environmental and food safety domains lags further behind (not least linked to later focus on these sectors, with data starting to be generated from 2024).

The evaluation finds that progressing from descriptive reporting to deeper interpretation and cross-sector linkage analysis is resource-intensive and skills-dependent. Stakeholders highlight that analytical capability, data cleaning burdens, and limited time and staffing constrain the depth and routine nature of analysis, even where data volumes have increased.

4.2.4 Progress in sharing of analysis – strong at international level, mixed nationally and weakly institutionalised at facility level

Sharing of analysed data strengthened most clearly at international level and also improved in most countries at national level from baseline in both human and animal health. Facility-level sharing of analysed data was more uneven, with strong examples in human health but limited progress in animal health. See Annex 4.2.4 for details

Facility-level[15] sharing. In human health, there is evidence – both quantitative and qualitative – of improved sharing of laboratory data with clinical teams and infection prevention and control/ antimicrobial stewardship structures, including through antibiograms and PPS outputs.[16] In animal health, local-level sharing is less clearly defined and more limited for compiled surveillance outputs, reflecting the structure of the sector and the design emphasis on active surveillance rather than practitioner-driven clinical decision support. Sharing of individual laboratory results under passive surveillance is routine, but mechanisms for feedback of compiled surveillance data to practitioners are less consistently evidenced.

National-level sharing. Results Framework reporting indicates increased sharing of surveillance outputs with national coordination mechanisms and wider dissemination processes in both sectors, including national conferences, which is confirmed in data collection within evaluation sample countries. This indicates expanding routinisation of national-level review and dissemination, though reporting does not fully distinguish sectoral patterns at this level.

International sharing. There is evidence of increased international reporting and data sharing through global platforms, and stakeholders from multilateral organisations describe the Fleming Fund's contribution as foundational to progress in global surveillance architecture, see Box 3 below. The Fund supported three key global platforms – WHO GLASS (human health AMR/AMU/AMC), FAO InFARM (animals and food systems), and WOAH ANIMUSE (antimicrobial

[15] See glossary for definition of 'facility'.

[16] As noted for data quality time and resources did not allow a more detailed assessment of the nature of these processes and how they support stewardship, e.g. in terms of frequency of updates / clinical feedback, denominator adequacy, stratification by ward or patient population, integration into prescribing guidelines.

use in animals). Reporting from Fund-supported countries increased over time, in terms of volume and scope. For example, all Fund-supported countries were reporting to ANIMUSE against Option 3 by 2024 (compared to three at baseline). Evidence has also been made available through the publication of a large number of peer-reviewed articles authored by Fleming Fund implementing partners and stakeholders.

Box 3 Illustrative examples of international sharing through global AMR platforms

WHO GLASS (human health AMR and AMC)

- Global participation increased from **25 countries in 2016 to 104 in 2023**, with Fund-supported countries rising from **eight (2018) to 18 (2023)**.
- Across evaluation case countries, reporting expanded substantially. For example, **over 139 hospitals in Indonesia** now report to GLASS, with evidence of reporting extending beyond Fund-supported sites.
- The **volume of AMR data reported** by Fund-supported countries also increased (from **34,156 isolates in 2018 to 148,389 in 2023**) and was maintained during COVID-19, in contrast to declines observed in other LMICs.
- Reporting of **AMU data** also expanded globally (73 countries in 2022 to 89 in 2023), though participation among Fund-supported countries remained uneven.

WOAH ANIMUSE (animal antimicrobial use)

- Participation reached **148 countries by 2024**, with **18 Fund-supported countries submitting data**.
- The **scope and quality of reporting improved** over time: at least nine Fund-supported countries progressed to more comprehensive reporting options, and **all reporting countries were using Option 3 by 2024** (compared to three at baseline).

FAO InFARM (animals and food systems)

- Launched in **2024**, with **49 countries reporting in its first year**.
- Participation from Fund-supported countries increased from **11 (2024) to 16 (2025)**, reflecting early uptake.

At the same time, the evaluation identified both missed opportunities with regards to sharing of country-level data[xix] and concerns that only relatively limited consolidated, readily accessible information is available for users seeking country-level AMR profiles. Constraints include data sovereignty and access arrangements, limited public availability of submitted datasets, and incomplete integration of Fund-supported data into global analyses. These factors limit the extent to which increased international reporting translates into wider analytical use.

4.2.5 Explanation of performance

The evaluation finds that the Fleming Fund made an important contribution to most of the key drivers of increased data generation and intermediate outcomes.[xx] Evidence from earlier evaluative work and Phase 2 evidence indicate that progress was driven by laboratory renovation and equipment provision, workforce investments, strengthened governance arrangements, and support to laboratory quality assurance systems. Other donors also contributed in some contexts, notably USAID, though the donor environment changed substantially during Phase 2.

Intermediate outcomes (data generated, analysed and shared) remain uneven across countries and sectors because they depend on both foundations and enabling conditions and require time to achieve. Growth in data volumes and reporting does not automatically produce national datasets that governments could readily use to make decisions where coverage, representativeness, data access, and analytical capacity remain constrained. In animal health,

uneven intermediate outcomes and weaker facility-level sharing are partly explained by the design emphasis on active surveillance to generate data relevant to public health priorities, which provides weaker incentives for practitioner-driven uptake within animal health systems.

Across countries, stronger performance in relation to analysis and data sharing has tended to be associated with a combination of practical and institutional enablers, including digitisation and automation (reducing data cleaning burdens), the availability of structured guidance and analytical tools, training and technical assistance, and the presence of functioning coordination mechanisms for review and dissemination. As highlighted in earlier analysis, these enablers appear to be most effective where they align with adequate analytical capacity and active demand for analysis within relevant governance or decision-making fora. Conversely, weaker performance is observed where analytical capability remains limited, interpretation and use place high demands on already constrained resources, roles and responsibilities for use are unclear, or mechanisms for review and action are weak or inactive. In some countries, these constraints have been compounded by the relatively short implementation windows for newer Phase 2 priorities, limiting the extent to which emerging analysis and sharing arrangements could become embedded.

4.3 Effectiveness: Longer-term outcomes

This section examines whether and how improvements in AMR surveillance - specifically data generated, analysed and shared - translated into longer-term outcomes in policy, practice, financing and governance. It recognises that such outcomes depend not only on the quality and availability of data, but on political priorities, institutional incentives and the formal and informal routes through which evidence reaches decision-makers, factors that sit largely outside any technical programme's direct control. See Annex 4.5.4 on governance and the enabling environment, and Annex 4.4 on sustainability determinants, which explore how these factors shape outcomes.

The following analysis builds on Section 4.2, which establishes that intermediate outcomes improved within the Fund's core sphere of influence, while also identifying limits to data usability and translation into national datasets that governments could readily use to inform decisions. The findings presented here draw on observable signals from programme-level data, international indicators, stakeholder perspectives and case-based evidence. They reflect what could be credibly assessed within the available evidence and timeframe, rather than a comprehensive account of all policy, practice or financing changes that may have occurred beyond what was visible through these sources.

The Fleming Fund has so far generated longer-term outcomes in terms of governance arrangements for action on AMR and in terms of establishing the conditions for changes in practice at supported sites. There is less evidence that intended results have been achieved in terms of downstream policy and regulation, or in terms of securing increased domestic resourcing for AMR surveillance.

Despite numerous examples of relevant policy and regulatory change during the period, this type of outcome remained limited, uneven and only weakly linked to improvements in surveillance. In most contexts, the Fund was not the decisive driver of action, partly because increases in AMR surveillance data were delivered with limited programme implementation remaining and were not primarily oriented towards national decision-making processes. There is evidence of the Fund's contribution to policy and regulatory changes in some countries. However AMR action in national policy agendas continued to be shaped primarily by political priorities, institutional mandates and external pressures, including compliance

with international standards – factors the programme's design and delivery mechanisms engaged with only to a limited extent and more recently (2023-2025).

The programme was more successful in developing changes in practice and attitudes, although these outcomes are early stage and emergent; they are not yet sufficiently embedded to demonstrate sustained or systematic change. Increased domestic resourcing for AMR surveillance was largely not achieved during the programme period, perhaps unsurprisingly given prevailing domestic fiscal contexts. The strongest and most consistent longer-term outcomes were in governance arrangements and the enabling environment – or conditions that make action possible (such as coordination, norms, platforms and communities of practice, political attention).

Longer-term outcomes were expected to strengthen during an intended third phase of support. While surveillance outputs expanded in volume and coverage towards the end of the programme, the evidence does not support the view that additional time alone would have generated substantial downstream change. Converting surveillance outputs into policy, practice or financing decisions depends on political prioritisation, incentives and enabling actions that the programme did not consistently or systematically address.

Across outcome domains, the pathways through which surveillance improvements might translate into longer-term outcomes differed markedly between human and animal health. In human health, lab-based AMR surveillance was more clearly embedded within health systems, with established institutional mandates, routines and clearer incentives to generate and use surveillance data. This created more consistent procedural signals of data use, even though translation into formal policy or regulatory action remained limited during the period.

In animal health, the relevance of lab-based AMR surveillance data to decision-making was more variable. Greater private-sector and market influence, more fragmented institutional responsibilities, and weaker routine demand for AMR-specific surveillance constrained the extent to which technical capacity alone translated into sustained outcomes. This does not imply lower overall potential for longer-term outcomes in animal health, but rather reflects the limits of a surveillance pathway centred on laboratories in this sector. Where the programme directly supported policy-oriented intervention, such as revisions to Standard Veterinary Treatment Guidelines and Essential Veterinary Medicines Lists under Phase 2, clearer policy signals emerged - helping to explain the increase in animal-health-related policy instruments observed in 2025 (see Figure 7 below).

Phase 2 sharpened expectations for outcome-level change, placing greater emphasis on the translation of surveillance outputs into decisions and action. However, Fund-level targets were intentionally not established for most longer-term outcomes, with implementers using international indicators and external systems for tracking higher-level change (in line with best practice).[17] In the absence of outcome targets, this section assesses longer-term outcome signals and explanatory factors across four domains – policy and regulation; practice and attitudes; national resourcing and financing; and governance and the enabling environment (see Figure 6 and Vol. 2 Annex 4.5 for examples).

It is important to acknowledge that the MA team, which implemented the significant proportion of Fund activity, especially at country level, was not contractually committed to delivering long-term outcomes beyond sustainability targets (Section 4.4). Observations about progress towards long-term outcomes should be interpreted in this context, and in view of the complex processes

[17] Best practice in terms of UK government guidance. At the time these were set out in the DFID Smart Rules (2020), subsequently replaced by the FCDO Programme Operating Framework https://www.gov.uk/government/publications/fcdo-programme-operating-framework. See also Vol 2, annex 3.4.3 for more details.

through which targeted changes take place, and bearing in mind the relative, lower priority attached to this agenda (as discussed in Section 4.5.4). The Theory of Change also anticipated that results at this level would occur across Phase 2 into an expected Phase 3.

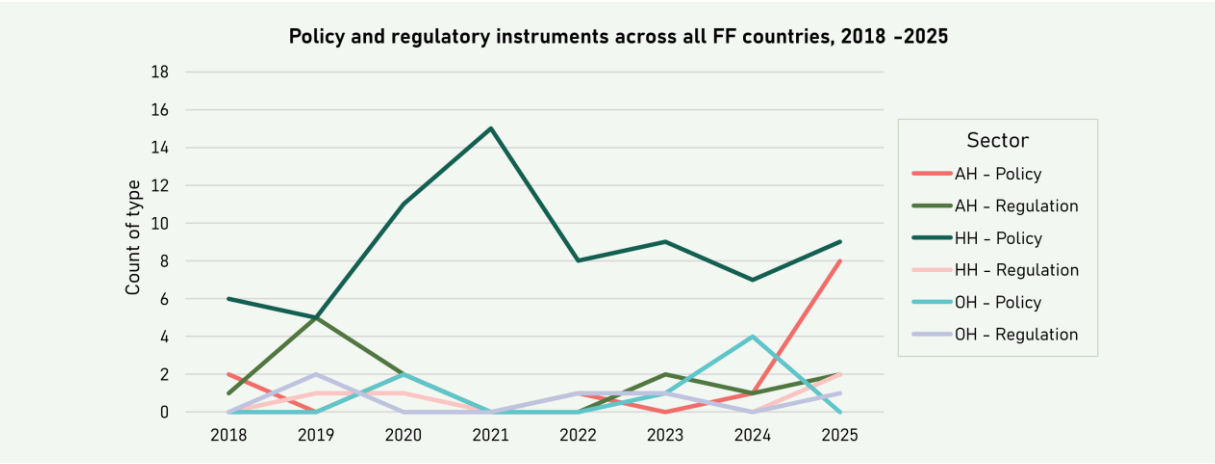
Figure 6 Types of longer-term outcome, with illustrative examples[xxi]

1 Policy	2 Practice	3 Resources	4 Governance
National policy & regulatory instruments showing countries taking action against AMR	**Evidence of changes in practice and attitudes amongst professionals working at facility level**	**Increased domestic financing for AMR surveillance through country budget allocation**	**Changes in the enabling environment, such as coordination, norms, platforms and communities of practice, political attention**
Example: Nepal prohibits import, distribution and manufacture of WHO 'not recommended' antimicrobials, 2023	**Example:** Timor Leste improves treatment of neonatal sepsis 2022	**Example:** Uganda Ministry of Health Quantification and Procurement Planning Unit (QPPU) work and Parliamentary Forum on AMR, 2025	**Example:** Kenya Fleming Fund Fellows documentary film to reframe the narrative, 2024

4.3.1 Policy and regulation change – limited, uneven and weakly attributable

Policy and regulatory change during the period was limited, uneven and weakly attributable to improvements in AMR surveillance. Across the 22 Fund-supported countries, there were only limited signs of concrete national action against AMR through new or updated policy or regulatory instruments, with such instruments more commonly observed in human health. Figure 7 presents counts of qualifying policy and regulatory instruments[18] across the implementation period. We note that National Action Plans (NAPs) and surveillance systems plans were excluded from this analysis because these primarily represent planning commitments rather than implementation-level policy or regulatory change. Further discussion on NAP alignment is included below, and in Vol 2 (Annex 4.5.1 and 4.5.4 (Box 12).

Figure 7 Policy and regulatory instruments, 2018–2025



[18] Official policies and legislation published by a country to address AMR, including Standard Treatment guidelines, Essential Medicines Lists and Infection & Prevention control manuals, or their equivalents.

Observed changes included new or updated standard treatment guidelines, revisions to essential medicines lists, and infection prevention and control instruments (see Box 4 for examples). Regulatory measures – such as restrictions on critically important antimicrobials – were observed in only a small number of settings.

Box 4 Illustrative examples of observed policy changes

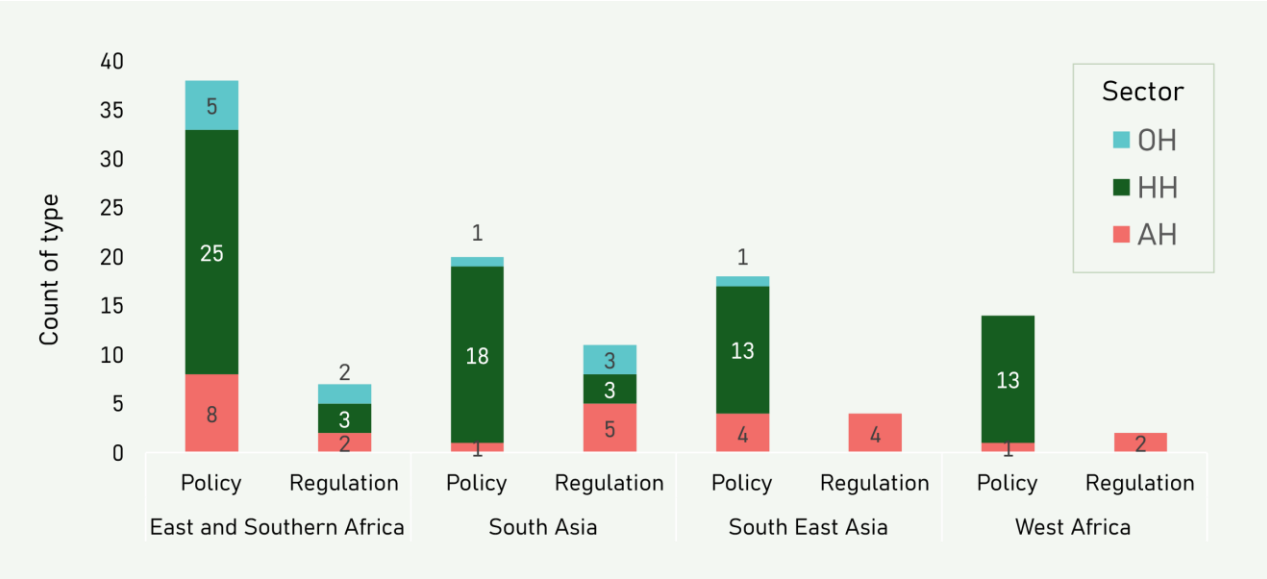
Drawing upon AMR surveillance data, the Department of Drug Administration in **Nepal** has taken key decisions: to restrict the dispensing of reserve group antibiotics to hospital pharmacies only; and to put mandatory red line markings on the packaging of all antimicrobial drugs. These decisions align with global standards to regulate and control the misuse of antibiotics.

Fleming Fund grantees supported the development of **Kenya**'s essential veterinary medicine list and treatment guidelines to help inform and promote the rational use of antibiotics.

More examples are available in Vol 2, Annex 4.5.5

Policy and regulatory developments were unevenly distributed across countries and regions. As shown in Figure 8, initiatives to take policy or regulatory action varied substantially across programme operating regions, with such instruments being approximately twice as common in East and Southern Africa as in other regions (see Annex 4.5.1).

Figure 8 Distribution of 2018–2025 Fleming Fund country instruments across regions, type and sector



Overall, patterns of policy and regulatory change do not show a consistent or systematic relationship with improvements in the availability, analysis or sharing of surveillance data (intermediate outcomes). In several contexts, reported increases in the 'use of local data' reflect procedural or reporting changes within coordination or accountability processes rather than the adoption of binding instruments for national action. Where policy or regulatory instruments do exist, evidence on their implementation or enforcement is often limited or falls outside what can be reliably observed through programme-level monitoring.

Taken together, this evidence does not support a conclusion that improved surveillance systematically drove policy or regulatory change during the period. Observed developments represent early or partial signals of influence. In human health, the Fund contributed indirectly in

some contexts, but rarely acted as the decisive driver of policy adoption, implementation or enforcement (see section 4.3.4 for discussion of drivers of data use). This contrasts with targeted animal health interventions under Phase 2, where policy change was pursued more directly as an output, explaining the recent increase in animal health policy instruments observed for 2025 in Figure 8.

4.3.2 Changes in practice and attitudes – early and emergent, concentrated at facility level

Changes in practice and attitudes were early stage and emergent, concentrated at facility level but not yet sufficiently embedded to demonstrate sustained or systematic change. Many countries report facility-level mechanisms intended to support the use of laboratory data, including infection prevention and control and antimicrobial stewardship structures, alongside improvements in how outputs such as antibiograms are packaged and disseminated. There is evidence that these mechanisms have strengthened awareness and created entry points for data use in some settings which is important.[19] However, sustained changes in prescribing behaviour are less consistently documented, and routine monitoring systems are not generally designed to capture changes in clinical practice or attitudes over time. As a result, while early signals of practice-level influence exist, confidence in the scale, consistency and durability of practice change remains limited.

Practice-change pathways are structurally different in animal health, where practitioner groups are more diffuse and where outputs from active surveillance may be less directly relevant to routine treatment decisions. The programme's approach to clinical engagement was therefore largely oriented to human health, with limited direct applicability to animal-health practitioner decision-making. While there are credible indications of early influence on practice and professional awareness, particularly at facility-level, the available evidence does not allow conclusions to be drawn about sustained, system-wide changes in behaviour. The Fund's contribution in this domain is therefore best understood as supportive and enabling, helping to create entry points for data use rather than driving consolidated change at scale. See Box 5 and Annex 4.5.2 for more details.

[19] Use of antibiograms and AMU data by clinicians, IPC teams, and stewardship committees is meaningful as it will engage clinicians and encourage laboratory use; surveillance only has value when it influences patient care

Box 5 Illustrative examples of observed practice changes

Fleming Fund support for AMR monitoring in the neonatal unit at Guido Valadares National Hospital (HNGV) in **Timor-Leste** led to the detection of a nosocomial outbreak of Phytobacter ursungii in 2025. Forty-two neonates were infected, with at least 10 fatalities. Outbreak response activities were conducted jointly by the clinical infectious diseases team and the Neo Sepsis team, with support from the Country Grantee. Subsequently, the number of cases reduced from 20 in April to five in May, and only one in June.

AMS committee meetings were held in 10 AMR surveillance sites in **Tanzania** during June 2025. Point prevalence survey findings were shared with hospital management. Committees noted the remarkable improvement in compliance with surgical prophylaxis guidelines compared to previous point prevalence survey reports.

Findings from clinical engagement assessments in Sindh and Gilgit Baltistan hospital in **Pakistan** (using AMR Stewardship, and Infection Prevention Control indicators) have been used to update facility SOPs, guide staff training and standardise clinical governance approaches

For further information on clinical engagement, see the MA's learning product: 'Antimicrobial resistance (AMR) and antimicrobial use (AMU) data use in clinical practice at Fleming Fund-supported human health sites (April 2026)', available on The Global Health Network (https://tghn.org/)

4.3.3 National resourcing and financing – largely not achieved

Increased domestic resourcing for AMR surveillance was only partially and unevenly achieved during the programme period, perhaps unsurprisingly given prevailing domestic fiscal contexts. Results Framework indicators show progress in the number of countries with costed plans, funding commitments and reported allocations, indicating movement towards resource mobilisation. However, these process indicators do not provide a clear basis for assessing the extent to which resources were actually increased, sustained or translated into consistent national-scale outcomes. Available evidence points to limited and uneven examples of tangible progress. This reflects both wider structural and fiscal constraints affecting AMR financing and the largely enabling and late-stage nature of economic and investment-case work undertaken during Phase 2 (see Section 4.2), which focused on developing tools, methods and analytic framing rather than securing routine uptake within national planning and budgeting processes. In most countries, given domestic fiscal constraints and levels of health financing, achieving domestically resourced AMR surveillance was a significant and in some cases unrealistic challenge. In countries with decentralised systems, variation in subnational capacity and prioritisation further constrained the translation of national commitments into consistent financing at scale.

Indicators related to implementation and financing of national AMR plans point to persistent constraints. Development of investment cases and economic framing to support mobilisation is uneven across countries, and evidence of their translation into routine national budgeting processes remains limited. Financing outcomes remain contingent on wider public financial management systems and political prioritisation processes that sit largely outside programme influence. Across countries, mechanisms for cost recovery, including user fees and national health insurance coverage for diagnostics, remain limited, under development, or insufficient to cover routine surveillance costs. Several stakeholders anticipated that, in the absence of replacement financing for reagents and consumables, routine testing and surveillance activity, including on-farm active surveillance, could decline towards pre-Fund levels in some settings – an issue that has direct implications for the sustainability of results, as discussed in Section 4.4. See Annex 4.5.3 for more details of country reporting on NAP financing.

Taken together, the evidence suggests that while the Fleming Fund strengthened surveillance capacity, it did not catalyse a sustained shift in how AMR surveillance is financed. Funding remains largely donor-dependent, with limited uptake of AMR surveillance costs into domestic budgeting processes (see Box 6 for examples). Engagement with finance authorities has been uneven and has not, to date, translated into durable financing decisions, leaving material risks to sustainability as external funding winds down. Where enabling work was undertaken, it has not yet translated into routine budgetary uptake, and any future financing outcomes will depend primarily on political and public financial management processes beyond the programme's direct influence. See Annex 4.5.3 for more details.

Box 6 Illustrative examples of observed changes funding decisions

Data collected in the five evaluation focus countries during 2025/26 countries (which is therefore illustrative rather than representative), provides some examples of increased domestic recognition of, and early steps towards financing, AMR surveillance:

- Indonesia: AMR included in the 2024–2029 National Development Plan and hospital accreditation standards, with partial federal funding for core functions, albeit still reliant on provincial and hospital resources.
- Nepal: Costed NAP endorsed, providing a basis for financing; small federal AMR budget lines are emerging, though still limited and requiring continued Ministry of Finance engagement.
- Nigeria: Adoption of costed NAP 2.0 has enabled federal agencies (e.g. NCDC, NAFDAC) to incorporate AMR into budgets, indicating growing institutionalisation, although funding release and prioritisation remain inconsistent.
- Kenya: AMR budget lines referenced in ministerial workplans, signalling policy intent, though not yet formalised or allocated in budgets.
- Pakistan: Launch of NAP 2 (January 2026) and prior PC1 funding demonstrate engagement with national planning processes, although future financing arrangements remain unclear.

Taken together, these examples suggest early but uneven progress towards integrating AMR surveillance into domestic planning and budgeting processes.

4.3.4 Governance and the enabling environment – the strongest and most consistent outcome-level contributions

The strongest and most consistent longer-term outcomes were in governance arrangements and the enabling environment – or conditions that make action possible (such as coordination, norms, platforms and communities of practice, political attention). Evidence from a peer-reviewed study using TrACSS data indicates that Fleming Fund support has been associated with changes since 2019 in how AMR data are used by public institutions, particularly within administrative decision-making, and to a lesser extent in informing policy, programme and regulatory processes.[xxii] Analysis for this report, updated to consider latest available TrACSS data from 2025, shows that Fleming Fund-supported countries reporting on use of data for decision-making shows stronger improvement than in comparable countries (see Annex 4.5.4 for details).

Stakeholders described increased political awareness and salience of AMR through sustained attention, evidence generation and engagement across sectors and levels. Global and regional normative frameworks and governance arrangements were strengthened through platforms, standards and guidance that countries reference in developing national systems. Networks and

communities of practice were reinforced across human and animal health, and to a lesser extent environmental domains, supporting knowledge exchange and professional cohesion.

Direct Grants contributed substantially to this enabling environment by supporting global and regional public goods - platforms, standards, tools, analytic capability, communities of practice and international coordination mechanisms - rather than uniform country-level outcomes. These contributions are most evident in support to global surveillance and data-sharing platforms (including GLASS, InFARM and ANIMUSE, highlighted in Section 4.2.4), strengthening technical capacity and progressive participation, and supporting coherence across sectors in global AMR fora. Contributions to political support for tackling AMR, including through the UN High Level Meeting in September 2024, were also significant; GRAM modelling appeared influential in this regard (see Annex 2.2 for details).

Programme support to countries was aligned with NAPs, with surveillance as one area of activity alongside others. While this helped frame and focus programme activities, it did not necessarily ensure the achievement of programmatic longer-term outcomes through translation of surveillance data into concrete action. NAPs remained relevant primarily as governance and coordination frameworks and for articulating resource needs, but there is limited evidence of resourcing or implementation at scale.[xxiii]

In contrast to other outcome domains, the evidence supports a clearer contribution claim in relation to governance arrangements and the conditions that make action possible. Here, the Fleming Fund plausibly contributed to strengthening norms, coordination mechanisms, platforms and professional networks, even where these effects do not translate directly into measurable policy, practice or financing outcomes within the programme timeframe.

4.3.5 Explanation of performance: Why outcomes progressed unevenly – evidence as support, not a trigger

This section explains variation in longer-term outcomes across the portfolio, drawing on evidence from Results Framework data, country case studies, TrACSS[xxiv] and key informant interviews. The evaluation applied case-based contribution analysis (described in Box 2 and Annex 4.5.1) to examine the Fund's contribution to drivers of policy and regulatory changes. These illustrative findings help clarify how change occurred in specific contexts, but should not be interpreted as a comprehensive assessment of contribution across the portfolio or as a test of the broader explanatory factors discussed below.

Box 7 Contribution analysis: policy and regulatory change in selected countries

The evaluation used contribution analysis to assess the plausibility of the Fund's contribution to selected policy and regulatory changes. The analysis focused on five case-study countries and is therefore **illustrative rather than representative** of contribution across the full portfolio. It is intended to demonstrate plausible pathways of influence where change was observed, not to quantify overall contribution or test all portfolio-level explanatory factors.

In four of the five cases, at least one policy or regulatory change between 2023 and 2025 was identified to which the Fleming Fund plausibly contributed, though this reflects the purposeful sampling strategy rather than the typical experience across all supported countries. The cases span both human and animal health regulation, with clearer contribution evident for **regulatory measures** than for broader policy change. Examples include: Kenya's National Antibiotic Use Guidelines on empiric treatment and surgical prophylaxis (2024); Indonesia's ban on quinolone use in animal production, supported by surveillance evidence and government engagement with the private sector; Nepal's prohibition of 103 antibiotic combinations across the supply chain; and Punjab's exclusion of resistant antimicrobials from veterinary procurement for 2024–25, based on provincial laboratory data.

Across these cases, change was most closely associated with shifts in **professional norms, understanding and capacity**, supported by strengthened networks and communities of practice. The Fund's contribution was most plausibly expressed through increasing recognition of AMR as a problem, improving technical and professional capability, and reinforcing shared standards. By contrast, there was limited evidence of direct contribution to political prioritisation or financing decisions, reflecting constraints beyond the programme's control. Further detail on individual cases, drivers of change and the Fund's contribution is provided in the annexes. Additional examples – particularly in animal health regulation – were identified but could not be fully triangulated and are treated as indicative only.

The factors outlined below synthesise portfolio-level conditions that have shaped progress in policy change, practice and resourcing. They are intended as explanatory lenses rather than a contribution framework. The Fleming Fund's contribution to specific policy and regulatory changes is examined separately through case-based contribution analysis (Box 7).

1. **Data follows action (evidence as a resource, not a trigger).** Surveillance data can strengthen salience and legitimacy for action, but rarely generates action on its own. Decisions are often driven by agenda alignment, leadership and incentives, with data used to justify, prioritise or refine choices rather than initiate them. As noted in Section 4.2, improvements in data generation do not consistently translate into accessible, national datasets that governments could readily use to make decisions, including where data ownership and sharing arrangements limit access.
2. **Political prioritisation and agenda alignment.** Policy and resourcing decisions are shaped by competing priorities and timing. Priorities for action were largely shaped by domestic policy processes, with countries pursuing approaches that reflected their own assessments, constraints and political economy, alongside, rather than being driven by, external support. Where AMR aligns with salient agendas, such as health security, food safety or quality of care, surveillance results were more likely to be taken up and inform discussion or action. Where such alignment was weaker, evidence was often acknowledged but did not translate into prioritised decisions, underscoring the importance of understanding and adapting to how partner countries framed the problem, rather than assuming that improved data alone would be sufficient to drive change

3. **Formal routes through which evidence feeds into policy and practice.** Translating evidence into action requires mechanisms to receive and interpret data, approve and implement policy instruments, and monitor change. Where these pathways are weak or fragmented across institutions, data may circulate without producing measurable change.
4. **Norms, standards and external reference points.** Global and regional norms, standards and platforms shape expectations about credible evidence and appropriate action. These reference points can influence behaviour more consistently than domestic datasets alone, helping explain stronger results in governance, standards, platforms and professional networks (Section 4.3.4).
5. **One Health as an aspiration and an operational constraint.** One Health is an important framing for AMR, but it remains challenging to operationalise. Different mandates, incentives, capacities and problem definitions across sectors limit integrated surveillance, joint analysis and coordinated decision-making. Implementation experience suggests a need to embed surveillance activity within existing sector programmes, rather than pursuing AMR as a cross-sector silo.[xxv] As a result of these constraints, One Health outcomes to date are more evident in coordination and information sharing than in identifying priorities and co-designing solutions.
6. **Timing, delivery-chain friction and administrative burden.** Several outcome-oriented shifts, notably clinical engagement and economic framing, were introduced late in Phase 2. This limited opportunities to institutionalise routines and observe mature effects. These constraints were compounded by delivery-chain obstacles, including repeated replanning and administrative requirements, which reduced adaptive capacity and constrained sequencing across grants.

Evidence across the portfolio indicates that data generation and use were shaped by both technical capacity and institutional factors such as prioritisation, incentives, mandates and demand. Capacity constraints were frequently binding in early Phase 1. Although Phase 2 introduced more explicit political economy and contextual analysis, this did not consistently translate into systematic reassessment of core assumptions about what was limiting routine surveillance practice as capacity improved. Even where such reassessment occurred, the scope to adjust delivery in response was constrained by the contractual model and the decision to end the Fund at short notice at the end of Phase 2, which limited flexibility to address incentive-, mandate- or demand-side constraints within existing grant structures.

4.4 Sustainability

Building on the outcome findings in Section 4.3, this section examines prospects for sustaining Fleming Fund-supported surveillance activities and capacities beyond the programme period, particularly in contexts where continuation depends on domestic prioritisation and recurrent financing.

In this evaluation, sustainability of AMR surveillance is understood not only in terms of retained infrastructure, equipment and skills, but also in terms of whether incentives, mandates and demand remain in place to sustain the routine production and use of surveillance data.

The findings in this section reflect what could be assessed, at the time of the evaluation, through available programme-level data, documentation and stakeholder perspectives regarding the likely continuation of both Fleming Fund-supported activities and capacities, and action towards AMR. They focus on prospective sustainability - what is likely to endure under current conditions - rather than a comprehensive account of all country-level commitments or future funding decisions that may emerge beyond the available evidence and timeframe.

Prospects for sustaining Fleming Fund-supported activities and capacities are generally limited but vary substantially across countries and sectors, particularly where continuation depends on high recurrent costs and ongoing Fund supported activities. Limited domestic financing in selected countries, the novelty of AMR surveillance, and structural features of the Fund's design meant that sustainability within a ten year horizon was unlikely. Prospects were further weakened by late transition planning, the decision to close the Fund at short notice at the end of Phase 2, and a more constrained external financing environment. As a result, while some elements of progress are likely to endure, routine AMR surveillance activities are unlikely to be sustained at current levels in many settings. The Fund's most enduring legacy is likely to lie in the human capital developed, the global networks fostered, and the sustained awareness generated.

4.4.1 Framing sustainability in the Fleming Fund context

This section assesses the likelihood that Fleming Fund-supported activities and capacities will be sustained beyond the programme period, what is likely to continue once the programme ends, and under what conditions, rather than reassessing achievements the Fund has supported, which is covered in Sections 4.2 and 4.3.

Sustainability was identified as an underpinning principle of the Fund from the outset, and Phase 2 placed increased emphasis on strengthening sustainability prospects. However, the Fund was conceived in a pre-COVID-19 context, and subsequent shocks, including compressed implementation timelines, shifts in national priorities, and a more constrained external financing environment, materially affected what could reasonably be sustained. These constraints are not unique to the Fleming Fund: long-standing global health programmes with decades of investment continue to face sustainability challenges, underscoring the difficulty of sustaining technically demanding systems within short funding horizons.

As noted earlier, greater emphasis was placed on sustainability in Phase 2 than during Phase 1. This included placement of sustainability-focused results in each Phase 2 Country Investment Strategy, charging country grantees with carrying out sustainability-focused political economy analyses (PEAs) during their inception periods, and development of Sustainability and Exit Plans (SEPs) by each grantee six months into their Phase 2 grants, based on those PEAs. As described in Vol 2, Annex 4.4 these efforts had mixed results.

Following confirmation in June 2025 that the programme would close at the end of Phase 2 (in March 2026), a range of actions were initiated, to protect gains and maximise the likelihood that priority elements of progress would endure. However, these efforts came too late to substantially alter overall sustainability prospects. These actions included roundtable discussions with partners, and the development of a small number of bespoke grants intended to extend or consolidate selected Fleming Fund objectives beyond programme closure. It is not yet possible to assess the effectiveness or durability of these interventions, given their late-stage introduction; but it is clear that expectations of what can be realistically achieved at this stage of the Fund's implementation should be limited. Nonetheless, these actions signal a clear commitment by DHSC to mitigate the impact of a short (six-to-nine-month) notice period before closure, protect gains where feasible, and codify learning through commissioned legacy and learning products. Further detail on these actions is provided in Annex 4.4.5.

4.4.2 What is likely to be sustained – and what is not

The overall sustainability judgement reflects on systematic differences in what the Fund was able to position for continuation, rather than on a single programme-wide sustainability outcome.

Routine surveillance activities

Sustaining routine AMR surveillance activities at current levels is unlikely in many settings without continued external support. Surveillance depends on recurrent financing for consumables, reagents, maintenance contracts and staff time. In a constrained fiscal environment, these costs are difficult to absorb into routine budgets, particularly where laboratory throughput remains low (as discussed in section 4.2 and see Vol 2, Annex 4.2.1 Box 7). Risks are most acute in animal health, where the Fleming Fund's predominant approach of high-cost, on-farm, active surveillance generates limited operational incentives for continuation (discussed in section 4.4.4). In addition to financing constraints, sustainability is undermined where surveillance data are not sufficiently required, incentivised or used within routine decision-making, reducing motivation to prioritise ongoing sample collection, testing and reporting.

Systems and capabilities

Some elements of laboratory and AMR surveillance system strengthening are more likely to endure even if routine surveillance activity declines. These include improvements to laboratory infrastructure, continued use of equipment that has lower maintenance costs, quality systems and data platforms that can continue to support broader diagnostic and laboratory functions. However, functionality remains contingent on maintenance, supply chains and staffing, and sustainability is therefore uneven.

Human capital and networks

Human capital represents one of the Fund's most durable legacies. The programme strengthened cohorts of trained professionals, professional networks and communities of practice across human, animal and, to a lesser extent, environmental health. These gains are likely to persist longer than surveillance activities themselves and can support continued AMR action through alternative pathways. This means sustainability of human resources is best understood as broader increased understanding, interest and capacity to address AMR, rather than guaranteeing sustained national surveillance activities. As evidenced in Vol. 2 (Section 4.4), in most countries capacity has not been fully institutionalised, with turnover and discontinuation of lab scientists and microbiologist posts posing a concrete risk to lab activities formerly supported by the Fund. Efforts supported by DHSC, such as inviting Fellows to join an alumni platform hosted by the Fleming Initiative[xxvi], promote continuity of networks and skill retention but can only partially mitigate these risks.

Governance and the enabling environment

Governance arrangements, norms and global platforms supported by the Fund show comparatively stronger sustainability prospects. These enabling outcomes – coordination mechanisms, shared standards, international platforms and political awareness – require lower recurrent financing and are embedded within multilateral and institutional structures, making them more resilient than country-financed delivery activities.

4.4.3 What will determine whether results last

Across countries, three interrelated factors most strongly shape sustainability prospects: resources, capacity and motivation.

Resources

As discussed in Section 4.3.3, persistent uncertainty around domestic financing creates significant sustainability risks for continuation of AMR surveillance. Sustainability risks therefore arise not from short-term delivery gaps, but from continued uncertainty around the availability of recurrent funding to maintain routine surveillance functions.

Evidence from country case studies and stakeholder interviews indicates that, in many settings, domestic resources are insufficient to cover the ongoing costs of laboratory-based surveillance,

particularly for reagents, consumables, equipment maintenance and staff retention. Where governments have committed resources, these tend to be partial, time-limited or reliant on external support, raising concerns about the continuity of routine testing and reporting once external financing ends. This is in spite of aforementioned efforts by the MA to steer end-of-programme dialogues with governments to consider the level of surveillance which could be both affordable and useful, if not at current levels.

In addition to domestic resource constraints, worsening external financing conditions have amplified sustainability risks. Reductions and reprioritisation in ODA have constrained the availability of follow-on external financing to support continuation of surveillance activities, increasing reliance on domestic resource mobilisation at a time when fiscal pressures remain acute in many countries. Where country ownership of surveillance functions was limited or incomplete, this shift in the financing environment has further reduced prospects for sustained continuation.

Capacity

The Fund has made substantial progress in building technical capacity, including through large-scale training and Fellowship schemes. These investments have strengthened awareness, skills and professional networks and are likely to endure longer than activity-level outputs (see Vol 2, Annex 2.2). However, capacity is not yet sufficiently institutionalised in many settings and remains vulnerable to staff turnover and declining system functionality if surveillance activities are not sustained.

Motivation and demand

Sustained use of surveillance systems depends on continued demand from clinicians, laboratories and policymakers. Phase 2 strengthened feedback loops, clinical engagement and awareness of AMR risks, contributing to increased motivation in some contexts. However, these gains are fragile: where surveillance activities are disrupted due to resource constraints, demand and engagement are likely to decline over time, weakening sustainability further. Where surveillance data continue to inform clinical practice, regulatory action or planning decisions, incentives to maintain routine data production and reporting are stronger; where they do not, demand weakens even where technical capacity remains.

4.4.4 Structural factors limiting sustainability

Beyond country-level determinants, several structural features of the Fund and its operating context constrain sustainability prospects.

Programme design, ambition and strategic clarity

The Fund's scale, ambition and technical focus, combined with a laboratory-centred model operating across 22 resource-constrained countries, made sustainability within a ten-year timeframe unlikely from the outset. Design choices such as reliance on automated equipment in human health and active, on-farm surveillance in animal health entailed high recurrent costs that were difficult to absorb through domestic financing systems.

In addition, no explicit, portfolio-level sustainability goals were agreed at the outset or during early implementation. While sustainability was articulated as a guiding principle and there was clear intention to strengthen long-term sustainability, expectations were not translated into a shared strategic view of what should realistically be sustained (for example, which functions, capabilities or outcomes), where, and over what timeframe. This limited the extent to which sustainability considerations could systematically shape programme design, prioritisation and adaptation across the portfolio, and constrained early dialogue with governments about long-term ownership and affordability.

Ultimately, the increased focus on data use from Phase 2 came too late and had limited scope to influence domestic resource mobilisation. This relates to the discussion on motivation and demand in Sections 4.4.3. and 4.3.Differentiation, ownership and tailoring of support

Sustainability prospects were further shaped by limited differentiation in the type and intensity of support provided across countries,[20] despite wide variation in starting conditions, fiscal space, institutional maturity and demand for surveillance outputs. While country grants were tailored to national contexts to some degree, the overall delivery model offered relatively limited scope to vary surveillance approaches, technology choices or ambition in line with countries' longer-term ability to finance, manage and institutionalise activities.

Levels of country ownership also varied across the portfolio. While the Fund took steps to secure government agreement to participate, this was not consistently accompanied by substantive commitments such as co-funding or explicit responsibility for recurrent costs. In the context of pressure to disburse funds, due diligence to test the depth of government ownership was limited in some cases. Memoranda of Understanding signed in several Asian countries represented a notable exception, providing a clearer basis for government engagement.

Animal health surveillance model

Sustainability challenges were most pronounced in animal health.[xxvii] While technical capacity and baseline data generation improved, routine continuation depended on incentives and mandates that were often weaker than in human health. The Fleming Fund predominantly pursued active, on-farm surveillance to detect resistance in zoonotic bacteria to antimicrobials considered critical for human health. While this approach was effective in generating baseline data and strengthening laboratory and workforce capacity, it offered limited operational or regulatory value to national animal health institutions in part due to underdeveloped understanding of shared One Health challenges. As a result, motivation to sustain routine data production was weak once external support ended, particularly given the high resource requirements and limited alignment with routine animal health mandates, decision-making processes and financing structures. More sustainable approaches, such as passive surveillance in some countries and integration with existing animal health systems, received greater emphasis in Phase 2 but were introduced too late to materially alter sustainability prospects within the programme timeframe. See Annex 2.5 and 4.4.3 for more detail.

Timing and late focus on sustainability

Although sustainability was an underpinning principle, in practice sustainability planning was completed with limited Phase 2 implementation remaining (12-18 months), constraining opportunities for national institutions to internalise roles, costs and responsibilities. Sustainability and exit planning took place under significant time pressure alongside competing close-out demands. Mixed messaging about programme continuation, followed by the decision to close the Fund at short notice at the end of Phase 2, weakened incentives for governments to commit resources or embed surveillance functions within domestic systems.

4.4.5 Overall sustainability assessment

Overall, sustainability prospects differ primarily by type of Fleming Fund intervention rather than by country context. Prospects are weakest where continuation depends on sustained domestic financing for high-cost surveillance activities, and stronger where the Fund contributed to human capital, governance arrangements and global public goods that are less resource-intensive and embedded beyond individual country systems. These patterns reflect the Fund's design choices

[20] Limited differentiation in terms of laboratory equipment provided, active surveillance in AH, supply-side interventions for data use, lack of co-financing agreements, funding a prescribed menu of activities (albeit that reporting or programme data organised against these may have obscured some variation).

and operating context, rather than weaknesses in delivery, and explain why sustainability outcomes vary substantially across result types and countries.

4.5 Value for money (VfM)

VfM findings draw on programme-level financial and monitoring data, case-study evidence and stakeholder perspectives, focusing on portfolio-level value within the programme timeframe rather than longer-term costs and benefits beyond what could be observed.

The Fleming Fund delivered mixed VfM. Performance was comparatively strong on cost control and aspects of delivery efficiency, reflecting delivery of a large and technically complex portfolio at scale with effective cost controls. VfM was weaker on effectiveness, equity and sustainability, particularly where benefits depended on sustained domestic financing, behaviour change and complex institutional reform. This pattern is consistent with ICAI portfolio-level findings on complex, multi-country technical assistance programmes. The decision to close the programme at the end of Phase 2 further weakened VfM by limiting the period over which intended benefits could be consolidated and realised. These patterns reflect the Fund's design choices, delivery model and operating context rather than weaknesses in grant-level delivery.

4.5.1 VfM framing and approach

VfM is assessed across the four standard dimensions of economy, efficiency, effectiveness and equity, covering cost control, delivery efficiency, achievement of results and who benefits.

VfM judgements are made at portfolio level, recognising variation by sector, country context and delivery modality (country, regional, strategic alignment and direct grants, plus Fleming Fellowships). The assessment draws on financial data, Results Framework reporting, country case studies and stakeholder perspectives, and is interpreted in light of the Fund's design parameters and the challenging implementation context described in Sections 2 and 4 (see Vol 2, Annex 4.6.5 for details). While the Fund adapted its intent over time, the delivery model and contracting arrangements provided limited scope for structured experimentation or redesign once implementation was underway, reflecting deliberate trade-offs between control, consistency and flexibility.

In VfM terms, part of the Fund's value sits at regional and global levels, not only within individual country results. Given the cross-border nature of AMR, investments that strengthen shared surveillance standards, analytical platforms and professional networks can improve preparedness and situational awareness beyond the countries directly supported. While such benefits are diffuse, accrue over long time horizons and are difficult to quantify, they form part of the value proposition alongside direct country-level outcomes.

Observed overheads and transaction costs should be interpreted in light of the Fund's delivery model and operating constraints. The MA-led approach delivered strong fiduciary control across a large, volatile multi-country portfolio, but, combined with extensive assurance requirements and limited flexibility to reprofile expenditure across years, generated higher transaction volumes and coordination costs. These costs reflect design trade-offs between control and flexibility in a constrained ODA operating environment (see Annex 4.6).

4.5.2 Economy

Economy is the strongest VfM dimension. Across both phases, the Fund demonstrated effective cost control and prudent financial management. Budget reviews and reprogramming processes

shifted resources away from indirect and management costs towards operational delivery, and many grants reported savings in excess of three per cent of initial budgets.

Centralised procurement of laboratory equipment generated measurable efficiencies, including significant savings (approximately £2 million) through bulk purchasing of automated technologies. These achievements reflect robust financial oversight and delivery systems, particularly within MA-managed grants.

However, strong economy performance coexisted with persistent affordability risks. Stakeholders consistently highlighted the high and volatile costs of imported consumables, long procurement lead times, and expensive equipment maintenance requirements. These factors contributed in some contexts to stock-outs and interruptions to testing activity, illustrating the limits of a standardised laboratory-strengthening model in resource-constrained settings.

4.5.3 Efficiency

Efficiency performance is mixed. The Fund delivered a substantial volume of activity across a large, multi-country portfolio, reflecting its technical ambition and scale. However, the delivery architecture created significant coordination and transaction costs that constrained efficiency.

The combination of country, regional, global and direct grants operating simultaneously in the same countries increased coordination demands. Delays in commissioning and contracting due to multiple rounds of review and sign-off for Phase 2 plans, particularly for regional grants intended to deliver Phase 2 strategic shifts, reduced effective implementation time and weakened sequencing between portfolio components.

Repeated reprogramming, no-cost extensions and redesign requirements, due to evolving priorities and external shocks such as COVID-19, Brexit, the Ukraine war and short-term UK spending reviews, further increased transaction costs and disrupted delivery. The scale and complexity of the portfolio meant that redesign and approval processes were often undertaken in parallel, amplifying opportunity costs. These efficiency constraints were driven less by grant-level performance than by portfolio-level design, coordination and approval processes, many of which reflected defensible assurance and risk-management choices. They therefore represent conscious design trade-offs shaped by system-level and contextual constraints, rather than delivery inefficiency or weak programme management. In some cases, however, these trade-offs had implications for equity (see Annex 2.7).

4.5.4 Effectiveness

From a VfM perspective, effectiveness was constrained relative to ambition. The Fund generated substantial outputs and intermediate outcomes, particularly in strengthening laboratory capacity, workforce skills and data generation. However, conversion of these achievements into sustained outcome-level change was limited and uneven.

It is clear that resources were managed effectively against what was contracted, but less effectively against later-stage outcome expectations. This reflects a design-level VfM constraint rather than an implementation failure. Contractual and performance management systems prioritised delivery of outputs and activities, while outcome ambitions evolved over time and were not consistently embedded in early contracting and accountability arrangements. At the same time, contractual arrangements limited the extent to which the programme could adapt to changing context, and key assumptions underpinning the design (see Section 2) were unrealistic.

The decision to close at the end of Phase 2 also constrained outcome-level results, particularly where benefits depended on longer-term institutional change, financing and behaviour change.

4.5.5 Equity

Equity is the weakest and least well-evidenced VfM dimension. Phase 1 did not include an explicit equity lens. Geographic equity was 'designed in' through the programme's focus on LMICs but direct beneficiaries were identified at the institutional level and LMIC and global populations were only considered indirect beneficiaries. Some isolated action to address equity was taken at activity level, mainly through recording gender balance of Fellows.

Gender and equity were given greater emphasis as a Phase 2 strategic shift, including through the Fellowship scheme and a specific Regional Grant (GEAR-UP). This produced some successes, such as the Fleming Fund-funded WHO Policy Brief on integrating gender in AMR NAPs[xxviii] (and the subsequent process in Uganda[xxix] to do so, supported by a GEAR-UP partner organisation). However, some opportunities were missed – for example to consistently embedded equity considerations in core programme delivery and capacity strengthening at all levels; as well as to encourage AMR surveillance systems to routinely collect and analyse sex- and age-disaggregated data or to publish first sex-disaggregated AMR burden estimates under the Global Burden of Disease study (due to data sparsity).

Progress on gender and equity therefore remains partial. Consistent with findings from the 2024 Progress Report, constraints reflected a combination of factors: weak initial conceptualisation, lack of guidance and targets/indicators; reliance on a dedicated Regional Grant with a delayed start and limited budget and mandate over other grantees; insufficient resourcing to mainstream gender and equity across all grantees; and contextual barriers, including low prioritisation by governments at country level and the gender composition of a limited pool of candidates for capacity strengthening.

Beyond gender, broader equity dimensions – notably geographic equity in site selection and equitable access to diagnostic services – were only partially addressed, in part due to the programme's emphasis on government ownership.

4.5.6 Overall VfM assessment

Taken together, the evaluation finds that, in most cases, VfM was strongest where benefits could be realised within the programme timeframe, through cost-conscious delivery, capacity building and enabling environment contributions. VfM was weaker where value depended on sustained domestic financing, complex institutional reform and equity mainstreaming that had been introduced late or remained under-resourced.

Importantly, weaker VfM in these areas reflects not only timing and resource constraints, but also key design choices that were poorly aligned with VfM considerations. The programme adopted a delivery model characterised by high recurrent costs, limited scope to vary surveillance approaches and technology choices across country contexts, and strong emphasis on laboratory-centred surveillance, which constrained affordability and sustainability in many settings. In addition, contracting and governance arrangements prioritised delivery against predefined outputs and offered limited scope to revisit or materially revise the core programme model once sustainability and VfM risks became apparent.

Investments in training, fellowships, professional networks, standards and global platforms represent among the more successful elements of the portfolio from a VfM perspective, as these generated system-wide benefits within the AMR surveillance ecosystem – including shared standards, platforms, skills and networks – and are more likely to endure than high-cost, delivery-intensive surveillance activities. Overall VfM patterns therefore reflect structural design and governance choices (including implementation arrangements), as well as operating context, rather than weaknesses in grant-level delivery.

5. Conclusions

This section draws together the evaluation's findings to provide an overall judgement on what the Fleming Fund achieved, how and why results were realised, and where limitations emerged over the past decade of investment. The conclusions support accountability to Parliament and UK taxpayers and are grounded in the evidence presented in Section 4. The evaluation assesses the Fleming Fund primarily against its original design logic and stated aims, while recognising that expectations around outcomes and sustainability evolved during implementation. A consolidated line-of-sight table mapping evidence sources to findings and conclusions is provided in Vol. II, Annex 5, alongside the detailed EQ-structured evidence in Annex 4.

Several of the constraints identified in this evaluation were already recognised during Phase 2 and informed early thinking about the direction and feasibility of potential future programming.

5.1 The Fund delivered strongly on its core intent

The Fleming Fund delivered strongly against what it was designed and contracted to do, establishing substantial AMR surveillance foundations and global public goods across a large, multi-country portfolio. (Grounded in findings on delivery against agreed outputs and intermediate outcomes; Sections 4.1 and 4.2 and Annex 4.1–4.2 (EQ1 Findings), with explicit evidence-to-conclusion mapping provided in Annex 5.)

The Fleming Fund made a reasonable strategic choice to prioritise AMR surveillance as its entry point and delivered strongly against this core intent. The programme substantially strengthened laboratory infrastructure, workforce capacity and surveillance foundations across a large, multi-country portfolio, and made a significant foundational contribution to global and regional AMR surveillance architecture, particularly in human health. This focus was appropriate given the scale of global evidence gaps, weak baseline systems in many partner countries, and the UK's technical strengths in laboratory systems, surveillance standards and analytics. Across both phases, the Fund delivered most strongly where objectives were clearly specified, technically bounded and aligned with its delivery model, with Phase 1 establishing substantial foundations in laboratory capacity, workforce development and surveillance outputs across 22 countries. These achievements represent the Fund's most significant results and are consistent with what it was designed and contracted to deliver.

5.2 The Fund generated important enabling and durable effects

Beyond its formal outputs, the Fund generated important enabling effects – skills, networks, norms and global platforms – that represent some of its most durable contributions and may outlast the programme. (Grounded in findings on enabling effects, regional and global contributions, and intermediate outcomes; Sections 4.2, 4.3 and 4.4.)

Phase 2 built on Phase 1 foundations by articulating stronger ambitions around data use, sustainability, equity and One Health integration. Although effective implementation time was short, Phase 2 contributed to learning, signalling intent and strengthening enabling conditions, rather than to consolidated outcome-level change.

Regional and global grants played a particularly important role in generating these durable contributions. By providing shared tools, training, analytical capability and platforms that individual countries could not have developed alone, they helped harmonise approaches, strengthen global reporting systems and reinforce professional capability across regions. These

enabling effects support AMR action through multiple pathways, particularly through strengthened human capital, professional networks and shared standards, even where routine national surveillance systems remain fragile.

5.3 Progress on deeper system-level change was limited and uneven

While surveillance capacity and data availability increased, progress on deeper, longer-term, system-level change beyond surveillance foundations – such as sustained policy uptake of surveillance outputs, routine data use, and domestic financing for surveillance functions – was limited and uneven. (Grounded in findings on data use, sustainability and outcome pathways; Sections 4.3 and 4.4.) Despite strong performance at output and foundation levels, the translation of improved surveillance into sustained policy, practice and financing outcomes remained inconsistent. Better data did not automatically lead to action, and the evaluation found no systematic relationship between increased data availability and subsequent policy or regulatory change.

Several design assumptions did not fully hold in practice. The programme implicitly assumed that strengthened laboratory capacity would naturally lead to data use, that countries would gradually absorb the recurrent costs of surveillance, and that cross-sector coordination would strengthen through shared structures. In practice, surveillance models that rely on relatively expensive automated platforms, proprietary consumables, and regular maintenance proved difficult for some countries to sustain once external support ended. While bacteriology testing necessarily requires reagents and consumables, lower-throughput laboratories can often operate more cost-effectively using simpler manual methods and non-proprietary supplies, reducing dependence on high-cost equipment and specialised service contracts. This suggests that surveillance system design should be aligned with local testing volumes, technical capacity, and recurrent financing to maximise long-term sustainability. Governance and coordination depended on political authority and incentives beyond programme influence, and effective implementation time was closer to seven years once delays and external shocks (e.g. COVID-19) were accounted for.

Design, scale and delivery choices also shaped these outcomes. The ten-year programme commitment, large country footprint and broadly standardised, laboratory-centred delivery model enabled delivery at scale but constrained depth of engagement, adaptability and differentiation. In non-human health sectors, the programme was oriented more towards understanding cross-sectoral transmission risks than towards influencing the drivers of antimicrobial overuse and misuse. As expectations expanded in Phase 2 to include data use, behaviour change and sustainability, these features limited the programme's ability to influence incentives, financing arrangements and routine practice, particularly in animal health systems.

As a result, while some countries made progress on governance, political awareness and selective policy adjustments, these changes were not systematic and did not follow a linear pathway from surveillance data to action across the portfolio as a whole.

5.4 Delivery constraints shaped what could realistically be achieved

Delivery performance was shaped less by grant-level failure than by structural constraints – including approval delays, repeated redesign and limited political leverage – which reduced effective implementation time and constrained outcomes. (Grounded in findings on delivery performance, approvals, replanning and adaptive capacity; Sections 4.1 and 4.5 and Annex 4.5 (EQ4 Findings), with methodological framing and limitations set out in Annex 3 and line-of-sight mapping in Annex 5.)

The programme delivered a complex multi-country portfolio, which is a notable operational achievement, but delivery was materially constrained by cumulative delays linked to approvals

and redesigns. These delays reduced practical implementation time and particularly constrained later Phase 2 priorities.

The delays largely reflected structural and procedural drivers rather than grant-level delivery failure, including layered approval processes, repeated redesign and compressed implementation timelines that reduced effective delivery time, particularly in Phase 2. These reflected the scale and complexity of the portfolio, layered approval and assurance processes, and repeated reprogramming as priorities evolved and external shocks (notably COVID-19 and supply-chain disruption) were absorbed. While some redesign was necessary and appropriate, the cumulative effect increased transaction costs and reduced available delivery time.

The delivery model also provided limited levers for sustained senior-level political engagement alongside technical delivery, particularly once delivery timelines narrowed and priorities expanded beyond technical surveillance towards data use and sustainability. Where progress depended on political prioritisation, cross-government incentives and resource allocation, programme influence was constrained, limiting what technical implementation alone could achieve.

5.5 Sustainability prospects and VfM are mixed

Overall sustainability prospects and VfM are mixed: performance was strongest where delivery could be directly controlled, but weaker where results depended on sustained domestic financing, institutional change and political prioritisation. (Grounded in findings on sustainability and value for money; Sections 4.4 and 4.5 and Annex 4.4 (EQ3 Findings) and Annex 4.6 (EQ6 VfM Findings), with relevant assumptions and constraints discussed in Annex 3 and consolidated lines of evidence set out in Annex 5.) Prospects for sustaining the Fleming Fund's country-level results are mixed and often weak, particularly where sustainability depends on continued domestic financing of routine, high-cost surveillance activities. Structural features of the programme's design – including its scale, technical ambition and reliance on cost-intensive, laboratory-centred surveillance models – meant that sustainability was challenging from the outset in many LMIC contexts, even where those models were effective in establishing surveillance capacity.

Late intensification of sustainability and transition planning, combined with the decision to close the Fund at short notice at the end of Phase 2 and a more constrained external financing environment, further weakened prospects for consolidation. As a result, many countries remain dependent on external funding for core surveillance functions, and routine surveillance activity at scale is unlikely to be sustained in the absence of additional support.

From a VfM perspective, performance was mixed. The Fund performed comparatively strongly on cost control and aspects of delivery efficiency, reflecting effective financial management and delivery of a technically complex portfolio at scale. VfM was weaker where benefits depended on sustained domestic financing, behaviour change and complex institutional reform. These patterns reflect the programme's design choices, delivery model and operating context, together with delivery-period dynamics such as late sequencing of sustainability planning and the decision to close the programme at short notice at the end of Phase 2, rather than failures in grant-level delivery.

Overall, the Fund's most durable legacy lies in people, networks, norms, standards and global public goods, rather than in fully institutionalised national surveillance systems. These are meaningful contributions, but they do not equate to sustained surveillance systems operating independently at scale.

6. Lessons

The lessons set out below are intended for designers, commissioners and senior implementers of AMR and other complex global health system programmes. Many of these lessons are not new: they reflect well-established insights from international development and global health practice, and the experience of the Fleming Fund is not unique within ODA-funded programmes in LMICs. Rather than presenting novel prescriptions, the lessons focus on operationalisation – how intent, design and delivery interacted in practice – particularly in a context of constrained implementation time and declining ODA, where clarity of purpose, affordability and sustainability are critical in shaping what can realistically be achieved. Each lesson draws on multiple strands of evidence presented across Sections 4.1–4.5, rather than on a single finding or case (see Annex 5 for a summary of how lessons build on findings and conclusions).

Lesson 1: Fix outcomes early; adapt pathways, not purpose

(Grounded in findings on programme design, contracting and outcome-level performance; Sections 4.1–4.3.)

Programmes should be explicit from the outset about the outcomes they expect to influence and maintain these throughout delivery. While outputs and delivery pathways can evolve, shifting outcome intent midstream weakens accountability, coherence and the ability to judge performance. If ambitions extend beyond delivery of technical foundations to sustained system-level change, this needs to be reflected clearly, and early, in programme objectives, scale and design.

Lesson 2: Protect strategic intent through the delivery chain

(Grounded in findings on procurement, contracting and delivery constraints; Sections 4.1 and 4.5.)

Strategic intent is often diluted as priorities move from business case to procurement, contracting and implementation. A small number of non-negotiable design principles, particularly relating to intended outcomes, sustainability and affordability, need to be embedded explicitly in procurement and contracting arrangements. Once delivery structures are in place, they shape behaviour more strongly than later strategic signals.

Lesson 3: Use grant-making where pathways are direct; design differently for complex system change

(Grounded in findings on delivery models, adaptability and system-level outcomes; Sections 4.2–4.4.)

Grant-based delivery is well suited to objectives with relatively direct, technically bounded output-to-outcome pathways. It is much less effective where success depends on complex system change, including shifts in incentives, institutional behaviour or political prioritisation. Where adaptive influence is central to success, delivery models, contracts and governance arrangements must explicitly allow for structured learning, experimentation and course correction within defined bounds.

Lesson 4: Do not assume linear pathways from evidence to action

(Grounded in findings on data use, governance and policy influence; Section 4.3.)

Improved availability of data and evidence rarely leads automatically to policy, practice or financing change. Evidence is one influence among many, interacting with political priorities, institutional pathways, leadership and timing. Where influencing action is an objective, programme design needs to engage deliberately with these wider drivers rather than assuming that better data alone will trigger change.

Lesson 5: Treat sustainability as a design imperative, not an end-stage activity

(Grounded in findings on sustainability, financing and institutionalisation; Section 4.4.)

Sustainability needs to be designed from the outset by working backwards from what must endure. Models that rely on high recurrent costs, low throughput or continued external support carry high risk in resource-constrained settings. Where sustainability depends on domestic financing and political prioritisation, early articulation of affordability and economic rationale is critical.

Lesson 6: Calibrate ambition to context and institutional capacity

(Grounded in cross-country and cross-sector findings on differentiation and VfM; Sections 4.2, 4.4 and 4.5.)

Effective programmes differentiate their ambition and delivery approach based on country context, sectoral maturity and institutional capacity. Lack of data is not always the binding constraint; challenges may lie in incentives, demand, feasibility or affordability. Large, standardised portfolios limit the scope for differentiation and increase the risk of applying high-cost models in settings where sustainability prospects are weak.

Lesson 7: Treat One Health as a design lens, not a delivery template

(Grounded in findings on One Health implementation across sectors; Sections 4.2 and 4.3.)

One Health is an essential framing for AMR, recognising that multiple sectors contribute to and are affected by AMR. Human, animal and environmental health systems differ markedly in mandates, incentives and capacity, all of which need to be understood in order to align priorities and co-design effective solutions. Effective One Health programming requires strong sector-specific foundations, selective integration and explicit choices about where integration adds value, rather than assuming convergence through shared structures alone.

Annexes

See Vol II for annexes, including:

1. Terms of Reference
2. Background: Fleming Fund design parameters
3. Methodology
4. Findings

4.1 Output-level reporting

4.2 EQ1: Generation of quality data, data analysis and data sharing, and EQ5: International sharing

4.3 EQ2: Alignment and coherence

4.4 EQ3: Sustainability

4.5 EQ4: Data use

4.6 EQ6 Value for money

5. Line of sight between findings and conclusions

---

[i] A seventh question relating to the sustainable integration of AMR burden calculation under the Global Burden of Disease (GBD) study initiated under the Fund-supported Global Research on Antimicrobial Resistance (GRAM) project is outside the agreed scope of this summative assessment and is therefore not covered in this report.

[ii] https://tghn.org/

[iii] https://academic.oup.com/cid

[iv] Surveillance is one component of an AMR response, as defined in the Global Action Plan on AMR (2015). See https://www.who.int/publications/i/item/9789241509763

[v] See Phase 1 summative report (Itad, 2023), plus reports from Fleming Fund Regional Grants: CAPTURA (https://captura.ivi.int/) and MAAP (https://aslm.org/wp-content/uploads/2024/05/MAAP-Final-brief.pdf)

[vi] This is a reflection of the scope and intent defined at the start of Phase 1. This does not imply that issues like use of surveillance data were not part of the thinking during Phase 1, as there is clear evidence that these formed important considerations. The distinction here is between formal intent for Phase 1 and Phase 2.

[vii] ICAI (2018) – CSSF aid spending reviewhttps://icai.independent.gov.uk/html-version/cssf/; ICAI (2017) – Rapid Review of the Cross-Government Prosperity Fund https://icai.independent.gov.uk/html-version/prosperity-fund/; ICAI (2019) – International Climate Finance: UK aid for low-carbon development: https://icai.independent.gov.uk/html-version/uk-climate-finance/

[viii] https://www.flemingfund.org/publications/fleming-fund-phase-i-evaluation/

[ix] Not published, this was an internal deliverable for DHSC

[x] Stated in MA contract, see ToR (p88): https://www.contractsfinder.service.gov.uk/notice/909bab4f-04b4-4ddb-a004-11ff8927538c?origin=SearchResults&p=1

[xi] Also discussed in Section 4.3.1 and in the Phase 1 summative report.

[xii] This does not suggest that reporting was not available for regional and global grants. We note that milestones and implementation against these were set for these grants. We refer here to Results Framework indicators and the extent to which it is feasible to identify regional and global contributions through these.

[xiii] https://icai.independent.gov.uk/html-version/icai-review-of-the-newton-fund/

[xiv] https://publications.parliament.uk/pa/cm5801/cmselect/cmintdev/260/26002.htm; https://icai.independent.gov.uk/html-version/rapid-review-of-the-prosperity-fund-approach-paper/

[xv] https://icai.independent.gov.uk/

[xvi] A series of reports produced by a consortium of international partners, including the Center for Global Development, the World Organisation for Animal Health, the Institute for Health Metrics and Evaluation, RAND Europe, Animal Industry Data, and the World Bank. https://www.cgdev.org/project/forecasting-fallout-amr-economic-impacts-antimicrobial-resistance-humans

[xvii] This assessment is based primarily on triangulation of key informant interviews, country case studies and document review, supplemented by Results Framework data where informative. Given the limited and uneven availability of outcome-level indicators, greater evidential weight is placed on qualitative evidence of how data were interpreted, discussed and acted upon in practice (seeAnnex 4.5).

[xviii] Also included in Annex 5.3.1 of the 2024 Progress Report.

[xix] Sharing of data from Fleming Fund-supported sites/countries to GBD/IHME to feed the GRAM project was not maximised, representing a missed opportunity. While there may have been data sovereignty issues limiting this, it could have been more systematically attempted, including encouraging countries in this regard. The MAAP and CAPTURA projects could have provided examples of how to do this successfully (see Annex 2.2 and MA Learning Product on 'Multi-country technical assistance approach for strengthening AMR surveillance in LMICs' for details on these initiatives).

[xx] Phase 1 summative report, p9

[xxi] https://www.flemingfund.org/publications/strengthening-amr-surveillance-stories-of-national-progress-in-africa-and-asia/

https://www.flemingfund.org/publications/capacity-building-case-study-timor-leste/

https://www.flemingfund.org/wp-content/uploads/a11fc0f88cea2c7c56e7c8302ebd8143.pdf

https://www.flemingfund.org/publications/demystifying-amr-in-kenya-the-real-life-impact/

[xxii] Drake, A., Sassoon, I., Armitage, J., Abbas, S., Maudling, R., Gupta-Wright, A., Serrano, A., Shafa, A., & Shorten, T. (2026). Country governance of antimicrobial resistance (AMR) surveillance: Observations on global progress and aid programme effectiveness using data from the Tracking AMR Country Self-Assessment Survey (TrACSS). *Globalization and Health*. https://doi.org/10.1186/s12992-025-01179-4

[xxiii] WHO, 2024 — *Antimicrobial resistance: accelerating national and global responses*, document A77/5 https://apps.who.int/gb/ebwha/pdf_files/WHA77/A77_5-en.pdf 'As at November 2023, 178 countries had developed multisectoral national action plans on antimicrobial resistance. However, in 2023 only 27%

of countries reported implementing their national action plans effectively and only 11% had allocated national budgets to do so.'

[xxiv] https://www.amrcountryprogress.org

[xxv] https://www.flemingfund.org/publications/optimising-amr-surveillance-in-animal-health/

[xxvi] https://www.fleminginitiative.org/post/fleming-fund-legacy

[xxvii] This assessment draws on synthesis of key informant interviews and country case-study evidence under EQ3 (Sustainability). Sustainability challenges in animal health were compounded where Fund-supported AMR surveillance was more closely aligned with human health priorities and international reporting requirements than with routine animal health decision-making. In such contexts, regulatory authority existed but did not consistently translate into demand for the specific AMR surveillance outputs generated, limiting incentives for sustained domestic financing once external support reduced (see Annex 4.3 and Annex 4.5).

[xxviii] https://www.who.int/publications/i/item/9789240097278

[xxix] https://gearupaction.org/moving-beyond-bugs-and-drugs-integrating-gender-and-equity-in-amr-programmes-in-uganda/